THE UNIVERSITY OF ALBERTA

AN ANALYSIS OF THE EFFECTIVENESS OF A METHODS IN-SERVICE PROGRAM FOR CERTAIN TEACHERS OF Mathematics 20

by

© WILLIAM NACIUK

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA

AUGUST, 1968

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance this thesis entitled "An Analysis of the Effectiveness of a Methods In-Service Program for Certain Teachers of Mathematics 20" submitted by William Naciuk in partial fulfillment of the requirements for the degree of Master of Education.

ABSTRACT

The study was designed to evaluate the potential of a methods in-service education program for effecting change in the teaching methods of certain teachers of Mathematics 20. The in-service program consisted of five two-hour discussion-lecture sessions held over a period of eight weeks. Printed materials outlining teaching activities in the "Mathematizing Mode" were given to the teachers. Parallels and contrasts were drawn with the "Expository Mode".

The two primary purposes were: (1) To determine change in classroom teaching practices of the participating teachers from pre- to during- to post- scores, and (2) To determine whether the extent to which teachers were able to vary their teaching methods were related to certain teacher characteristics. Another purpose was to investigate and develop instruments and procedures through which measures of teaching methods could be obtained.

The research design utilized one experimental group of seven teachers, each of whom was required to teach one class with the "Mathematizing Mode" and another class with the "Expository Mode". Classes were assigned to the two treatment methods in such a way as to balance, as far as possible, the achievement level of the students in each treatment group.

For the purpose of investigating the extent to which teachers were able to change their teaching practices in the direction of the two experimental teaching models, two instruments were developed by the writer from a review of related research and current literature. The Observer Rating Scale was used by the investigator to obtain

measures of teaching methods during two live observations of each teacher's lessons in each treatment class before the in-service, during the experimental period, and after the experiment. The Student Inventory, which was designed to parallel the first instrument, was administered to all students during each of the above periods to elicit student opinions on the nature of the classroom activities. The pre-, during-, and post- scores so obtained from each instrument were translated into ranks. Rank orders were analyzed for differences Between Teachers and Within Teachers for methods.

The findings gave evidence that a methods in-service program has potential in effecting change in the teaching methods of certain teachers. Furthermore, the results indicated, although not significantly, that the extent to which teachers were able to teach the two treatment classes using two different methods was positively related to permissiveness and negatively related to the number of degrees held and the number of times the teacher had previously taught the course. During the experimental period six of the seven teachers were found to be teaching the two treatment classes with methods significantly different from those they used before the experiment. The methods used by one teacher during this period were not significantly different between the treatment classes. After the experimental period, four of the seven teachers were found to be teaching the two treatment classes with methods significantly different from those they used before the in-service treatment. In general, it would appear that not all teachers can equally translate theoretical teaching methods into classroom practice, nor do all teachers sustain to the same degree newly acquired teaching practices.

ACKNOWLEDGEMENTS

The writer wishes to acknowledge his indebtedness to Dr. T. E. Kieren, Chairman of the thesis committee, and to Dr. S. E. Sigurdson. The planning of the study, the conduct of the in-service program, and the writing of the thesis was made possible by their willing contributions, their patience, advice, and criticism. The writer is grateful for the valuable suggestions and constructive criticism given by Dr. L. D. Nelson, member of the thesis committee.

The writer is also indebted to Dr. D. B. Harrison who offered valuable assistance in the writing of the Fortran program to compile the data, and who offered suggestions for its treatment. Appreciation is also expressed to members of the Division of Educational Research Services who so willingly assisted in the preparation of the IBM Data cards.

Appreciation is most gratefully expressed to those teachers and their students who so kindly gave of their time and effort to take part in this investigation. Without their co-operation this investigation would not have been possible.

Appreciation is also expressed to Miss Sharon Humphreys for so willingly and cheerfully typing the many drafts of the instruments and the thesis.

And last, but not least, the writer wishes to express his gratitude to his wife, Jean, and to his three sons, David, Terry, and Gregory for being so patient and understanding throughout the progress of the study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM IN PERSPECTIVE

I. INTRODUCTION

In planning a lesson, the teacher must consider the two-fold problem: first, what are the proper objectives of the lesson, and second, what are the most effective instructional procedures through which the objectives may be achieved?

The first problem, although being fundamental, is basically philosophical. It cannot be ignored, however, since the appropriateness of an instructional method can be properly assessed only in terms of the purposes for which it is used.

The second problem, what are the most effective instructional procedures through which the objectives may be achieved, implies that the teacher should possess an extensive repertoire of instructional techniques and methods from which he may draw. Joyce (1966) underlined the importance of this when he wrote:

> A teacher who can purposefully exhibit a wide range of teaching styles is potentially able to accomplish more than a teacher whose repertoire is relatively limited.[1]

---

[1] Bruce R. Joyce and Richard E. Hodges, "Instructional Flexibility Training", The Journal of Teacher Education, Vol. XVII, (Winter, 1966), p. 409.

The teacher then, should choose from his repertoire those methods of instruction which will help him meet the objectives of instruction most effectively, be flexible and proficient in the use of the methods chosen, and constantly review and revise his repertoire.

II. BACKGROUND TO THE PROBLEM

In the last decade there has been a great ferment at all levels of mathematics curricula to produce new instructional materials stressing new methods as well as new content. A perusal of curriculum guides prepared for Alberta teachers of the new mathematics programs indicates that a variety of methods should be employed in the teaching of school mathematics. One of the major problems confronting the teachers of the new materials appears to be that of establishing a new and enlarged repertoire of methods from which they may draw for their daily instructional purposes.

This may require some sort of training or retraining of many of the teachers now teaching mathematics courses. One commonly used technique for retraining teachers is the in-service education program. It has the advantage of not requiring teachers to leave their teaching positions for prolonged periods of time.

Since attitudes and behavior patterns are persistent[2], it may be that teacher behaviors cannot be effectively changed through an in-service program. Furthermore, certain teacher characteristics may be more conducive to favourable change than others. Most of the research described in the literature is concerned with the effectiveness of various methods of instruction from the student achievement point of view rather than assessing the in-service program's effectiveness in changing teacher classroom behavior.

III. THE PROBLEM

Statement of the Problem

The problem of this study was two-fold. First, the study was designed to evaluate the relative effectiveness of a methods in-service education program for effecting change in the established teaching behaviors of certain high school mathematics twenty teachers. Second, it was designed to research and develop procedures and instruments through which the effectiveness of the in-service program could be measured objectively.

Context of the Problem

Each of the seven teachers involved in the experiment were required to attend five methods in-service sessions during which

---

[2]Ned A. Flanders, "Teacher Influence, Pupil Attitudes, and Achievement", Studying Teaching, James Raths et al, editors (Englewood Cliffs: Prentice-Hall, Inc., 1967), p. 45.

specific activities peculiar to the Mathematizing and Expository modes of teaching were presented and explained. Each teacher was then required to teach two units on the quadratic function of Mathematics 20 to one class using the Mathematizing Mode and to the other class using the Expository Mode. The two methods are defined briefly as follows:

Mathematizing Mode - is a method in which there is an initial period of uninhibited exploration of the problem situation on the part of the students. This is followed by a "brain-storming session" of hypotheses formulation in which the teacher acts as the moderator. Questions and problems are initiated by the teacher which enable the students to test out their hypotheses. The final step consists of "summing up" of the mathematical principles evolved in the preceding sections.

Expository Mode - is a method in which rules and theories are presented and explained by the teacher. Examples are given the class in which the rules or theories are clarified or applied. Students are questioned on the understanding of direct items of the topic. Assignments are given to provide further practice with the application of the rules or concepts discussed in class.

Hypotheses

This study attempted to provide information that could be used as a basis for answering the following general questions:

1. Do certain teachers of high school mathematics now teaching use a variety of methods in their instructional lessons or are the methods used essentially the same?
2. Do certain teachers of high school mathematics who are more flexible in their methods possess certain identifiable characteristics?
3. Can the established teaching methods of certain high school teachers of mathematics be effectively changed through an in-service program that is specifically designed to prepare teachers with particular methods?

Specifically, the hypotheses tested were:

1. There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes before the in-service treatment.*
2. There is no significant difference between teachers methods ranks as indicated by the Teachers Category Index scores before the in-service treatment.
3. There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes during the experiment.

---

*For definitions of <u>teachers' methods ranks</u> and <u>Teacher Category Index scores</u>, see pages 79 and 57 respectively.

4. There is no significant difference between teachers methods ranks as indicated by the Teachers Category Index scores during the experiment.
5. During the experiment, there is no significant difference between ranks of teachers when their ranks are determined by:
   a) age and flexibility of methods
   b) years of teaching experience and flexibility of methods
   c) number of times they had taught mathematics twenty and flexibility of methods
   d) number of university mathematics courses taken and flexibility of methods
   e) number of university degrees held and flexibility of methods
   f) percentile scores obtained on the Minnesota Teacher Attitude Inventory and flexibility of methods
6. There is no significant difference in the ranks of flexibility of methods between sexes during the experiment.
7. There is no significant difference in a teacher's methods for classes before the experiment and after the experiment.
8. There is no significant difference between teachers' methods for classes after the experiment.

Delimitations of the Problem

Since each of the teachers involved in the experiment was required to be teaching at least two classes of Mathematics Twenty, and each was also required to attend five in-service sessions out of school, only seven teachers of the Edmonton Public School System were able to take part. As the in-service program was carried out during the months of January and February, with the experiment ending in April, certain instructional practices and procedures had been established in the classes prior to the experiment. The investigation was further limited to the length and type of the in-service program of the experiment. Finally, the extent to which the instruments used gave adequate measures of instructional methods was a limitation.

IV. SIGNIFICANCE OF THE PROBLEM

With the introduction of the new mathematics programs there has been increased pressure on the teachers to employ a variety of teaching methods in their instruction. An abundance of evidence exists in the literature supporting the contention that many high school teachers of mathematics do not have the educational background to effectively employ different methods. Dossett (1964) indicated the need for a positive approach to the problem when she

wrote, "What is needed is not further evidence of teachers' lack but ways in which the situation may be remedied.[3]

Since it is impractical to withdraw teachers from their duties until they have retrained themselves, school authorities, educators, and teachers must investigate various forms of in-service educational programs' potential for effectively helping teachers to establish a repertoire of methods. Literature is virtually overflowing with research evidence on the relative effectiveness or ineffectiveness of various methods of instruction on student achievement, but there is a paucity of evidence on the effectiveness of in-service programs on changing teacher classroom behavior patterns especially for teachers of high school mathematics.

Houston (1961) underlines the significance of the problem in the following way:

> Studies are needed which investigate the changes brought about in the classroom as a result of in-service education programs ... such studies must be forthcoming if continued effort devoted to in-service education is to make the greatest possible contribution to the continued growth of teachers.[4]

---

[3]Mildred J. Dossett, "An Analysis of the Effectiveness of the Workshop as an In-Service Means for Improving Mathematical Understanding of Elementary School Teachers", Unpublished Doctoral Dissertation, (Michigan State University, 1964), p. 11, (microfilm)

[4]W. Robert Houston, Claude Boyd, and M. Vere DeVault, "An In-Service Program for Intermediate Grade Teachers", <u>Arithmetic Teacher</u>, Vol. 8, (February, 1961), p. 65.

This investigation is significant to the extent that it attempts to add objective data to the stated problem and to the whole problem of techniques needed for such an investigation.

V. EXPERIMENTAL PROCEDURE

The investigation reported here is one part of a much more complete group project. Nine teachers of Mathematics 20 in the Edmonton Public School System volunteered to take part in the project. Since each of the teachers was required to teach at least two classes of Mathematics 20 for the purposes of this study, this study had seven teachers participating.

Each teacher attended five two-hour in-service sessions in January and February, 1968. The in-service sessions consisted of lectures and discussions on the Mathematizing and Expository methods of teaching quadratic functions (Chapters Nine and Ten) of the Mathematics 20 course. Participants were given hand-outs prepared by Johnston explaining the activities to be carried out in teaching these chapters using the Mathematizing Mode.[5]

During the months of February and March, each teacher was required to teach one class using the Mathematizing Mode, while the other class was to be taught with the Expository Mode. Classes were assigned randomly to the methods.

---

[5]Ross Johnston, "The Mathematizing Mode," (A Working Paper, University of Alberta, Edmonton, 1968).

Data on teacher characteristics were obtained from the Teacher Characteristics Questionnaire and from the Minnesota Teacher Attitude Inventory which each teacher completed. Another questionnaire was completed by the teachers on their attitudes to and impressions of the whole experiment. This questionnaire was completed in June.

Data for each of the teacher's behavior patterns (methods of instruction) were collected through the application of two instruments prepared by the investigator designed to give relative measures of methods on an Expository - Mathematizing continuum. The Student Inventory of Teaching Behavior was filled out by each student before, during, and after the experiment. The Observer Rating Scale was used by the investigator while observing the teachers operating within each class. Two observations were carried out on each class before, during, and after the experiment.

The findings of the experiment are based upon statistical analyses of data obtained from the instruments as indicated above.

VI. OUTLINE OF THE REPORT

The present chapter is an introduction to and a preview of the study. Chapter II consists of a review of the literature that bears on the study being reported. Chapter III contains a detailed description of the design of the experiment, the education

in-service program, and the statistical procedures used in analyzing the data. The rationale, preparation of the investigator designed instruments, and the pilot study are reported in Chapter IV. The results from the statistical analyses and the teacher questionnaire are presented in Chapter V. Chapter VI contains a summary of the investigation together with conclusions, limitations, and recommendations for further research.

CHAPTER II

REVIEW OF THE LITERATURE

This study is concerned with: (1) investigating the extent to which teachers are able to adjust their teaching methods as a result of a methods in-service program; (2) relating this adjustment to certain selected teacher characteristics; (3) investigating and developing a technique by which the extent of adjustment of methods may be measured.

The literature reviewed is divided into two sections. The first section contains two parts; first, a review of the literature concerned with the state of methods in-service programs and second, a summary of research studies which have attempted to evaluate the relative effectiveness of in-service educational programs in producing changes in the teachers' instructional patterns. The second section is devoted to the review of the literature pertaining to the problem of analyzing and evaluating classroom interaction patterns for methods of instruction.

I. REVIEW OF LITERATURE ON METHODS IN-SERVICE EDUCATION PROGRAMS

Both teachers and educators are concerned with ways in which the quality of classroom instruction may be improved. Flanders (1963), in discussing teacher behavior and in-service programs, identified

introduction of new programs, programs built around some theme such as "individual study", and methods or flexibility training of teachers as three ways in which attempts have been made at accomplishing this goal.[1] In reflecting on summer institutes for teachers, Fenton and Ford (1966) made the observation that a number of techniques are useful in the classroom, depending on the objectives and materials chosen for the day's lesson.[2] They further stated that many objectives in the cognitive domain can be best attained by methods different from those needed to attain objectives in the affective domain. Hawkins (1966) stated that the essence of good teaching is not a simple one: "More than likely it is a subtle mapping across conspicuous variations in the local culture and individual style."[3] Others, such as Beberman (1959) and Davis (1966), stated that to teach their programs effectively, teachers need training in methods as well as in the content and structure of their courses.[4] It appears that

---

[1]Ned A. Flanders, "Teacher Behavior and In-Service Programs," Educational Leadership, Vol. 21, (October, 1963), p. 25.

[2]Edwin Fenton and Richard B. Ford, "Organizing Summer Institutes for Teachers: Some Reflections," The Journal of Teacher Education, Vol. XVII, (Spring, 1966), pp. 53-59.

[3]David Hawkins, "Learning the Unteachable," in Learning by Discovery: A Critical Appraisal. L. S. Shulman and E. R. Kieslar (eds.), (Chicago: Rand McNally and Company, 1966), p. 5.

[4]Max Beberman, "Improving High School Mathematics Teaching," Educational Leadership, Vol. XVII, (December, 1959), pp. 162-166; and

Robert B. Davis, "Discovery in the Teaching of Mathematics," in Learning by Discovery: A Critical Appraisal. L. S. Shulman and E. R. Kieslar (eds.) (Chicago: Rand McNally and Company, 1966), pp. 114-126.

to teach effectively toward the objectives of the new programs, teachers must have a clear understanding of the content and structure of the subject matter and must judiciously choose techniques or methods which will be most effective in helping the students reach the objectives. Clearly, this cannot be done on the basis of pre-service training alone, but can result only through in-service training programs carried out on a continuous basis.

The State of Methods In-Service Programs

In their review of the literature on pre-service and in-service education of teachers, Denemark and MacDonald (1967) concluded that: "Research on in-service education, considered as integral part of teacher education, was disappointingly scanty."[5] They further stated that it appeared from the review of the literature that the large grants for teacher education have been given for program development and not for theory development or research activity.

The Manual for the Preparation of Proposals for institutes sponsored by the National Defense Education Act under Title XI, states in part, that: "In every instance subject matter is to be emphasized,"[6]

---

[5]George W. Denemark and James B. MacDonald, "Pre-Service and In-Service Education of Teachers," Review of Educational Research, Vol. XXXVII, (June, 1967), pp. 240-241.

[6]Manual for the Preparation of Proposals, cited in Edwin Fenton and Richard B. Ford, op. cit., p. 53.

although, "where applicable, instruction in the use of the new educational techniques, media, and materials should be part of the program."[7]

Heideman (1962), in the background to his study of the participants of National Science Foundation Academic Year Institutes, stated that the primary objective of that program is to upgrade the academic competence of the teachers presently trained in science and mathematics. A secondary objective of the program, he states, is to allow for maximal development of modern teaching methods.[8]

In Alberta, two series of in-service education programs have been available to high school mathematics teachers during the past several years. One set of programs has been sponsored by the Faculty of Education through the Department of Extension, University of Alberta, while the other set by the Mathematics Specialist Council of the Alberta Teachers Association. A perusal of the brochures of each of these programs indicates that the primary objective of the programs has been to familiarize the participants with the textbook content

---

[7]<u>Ibid</u>.

[8]Robert G. Heideman, "National Science Foundation Academic Year Institutes for Secondary School Teachers of Science and Mathematics Held at the University of Wisconsin, 1956-1959: An Evaluation of the Background, Training, Placement, and Occupational Mobility of the Participants," (Unpublished Doctoral Dissertation, (University of Wisconsin, 1962), p. 3 (microfilmed).

and mathematical concepts of the new mathematical programs. A secondary purpose was to acquaint teachers with new methods of teaching and classroom use of the textbooks.[9]

Literature reveals that the greatest emphasis of in-service education programs has been in improving teachers' competence from the point of view of content and structure of the new mathematics programs rather than from the point of view of new methods of instruction. The latter is receiving some attention in a secondary sense.

Review of the Literature on Methods In-Service Programs

A review of the literature on in-service education programs whose main objective is to enumerate a variety of teaching techniques or strategies which reflect particular methods of teaching revealed that such in-service programs are few and far between. Furthermore, the effectiveness of in-service programs per se in many cases was evaluated from the point of view of changes in teacher attitudes and student achievement rather than from the point of view of sustained changes in teacher classroom behavior patterns.

As a secondary study of an experiment designed to compare the Cuisenaire and Dienes methods with others, Williams (1966) observed the extent to which the participating teachers used the two methods

---

[9] Department of Extension, "Continuing Professional Education for Teachers," (Edmonton: University of Alberta, 1967).

and mathematical concepts of the new mathematical programs. A secondary purpose was to acquaint teachers with new methods of teaching and classroom use of the textbooks.[9]

Literature reveals that the greatest emphasis of in-service education programs has been in improving teachers' competence from the point of view of content and structure of the new mathematics programs rather than from the point of view of new methods of instruction. The latter is receiving some attention in a secondary sense.

<u>Review of the Literature on Methods In-Service Programs</u>

A review of the literature on in-service education programs whose main objective is to enumerate a variety of teaching techniques or strategies which reflect particular methods of teaching revealed that such in-service programs are few and far between. Furthermore, the effectiveness of in-service programs per se in many cases was evaluated from the point of view of changes in teacher attitudes and student achievement rather than from the point of view of sustained changes in teacher classroom behavior patterns.

As a secondary study of an experiment designed to compare the Cuisenaire and Dienes methods with others, Williams (1966) observed the extent to which the participating teachers used the two methods

[9]Department of Extension, "Continuing Professional Education for Teachers," (Edmonton: University of Alberta, 1967).

after the experimental period. He found that teachers tended to revert to familiar methods. He concluded that ways need to be found to implement changes; that is, getting teachers to exchange the familiar for the new methods, and getting teachers to sustain these changes.[10]

Roscoe (1958) compared the extent to which two workshops held at the Alabama Polytechnic Institute met the needs of the twenty-five participating teachers with varied interests and characteristics.[11] His conclusions, based on the analysis of teacher opinions as reflected on the investigator prepared questionnaire, were that all teachers varied their teaching methods as a result of the workshop in-service. He found no significant difference in the effectiveness of the in-service program with teachers trained to teach mathematics and teachers trained to teach in other areas. Significant differences in the effectiveness of the in-service program were found between teachers having ten to fifteen years of experience and teachers having less than ten or more than fifteen years of experience. No indication as to the direction of the difference was stated.

---

[10] J. D. Williams, "'Method-Reversion': The Problem of Sustaining Changes in Teacher-Behavior," Educational Research, Vol. VIII, (February, 1966), pp. 128-133.

[11] Douglas K. Roscoe, "An Analysis of Two Mathematics Workshops for Teachers and Outcomes as Reflected in Participating Schools," Unpublished Doctoral Dissertation, (Alabama Polytechnic Institute, 1958).

Flanders (1963) conducted a project in which fifty-one junior high school teachers participated in two different types of in-service training programs, each lasting nine weeks. The purpose of the in-service training was to increase the flexibility of teacher influence and to increase the use of those teacher behaviors which support pupil participation in the classroom learning activities.[12] One conclusion he drew from the results of the project was that those teachers most active in the training program made most progress in the direction of the objectives of the program.

Shaver (1961) investigated, through systematic observations, the ability of four experimental teachers of the Harvard Social Sciences Research Project to conform to two different teaching styles.[13] Styles of teaching were explained to the experimental teachers by the investigator prior to the experimental period. Lesson plans based on each teaching model were prepared for the teachers. Although differences were found among the teachers between and within the styles, no correlation between these differences and teacher characteristics was reported. Shaver does suggest, however, that studies should be

---

[12]Flanders, *op. cit.*, p. 29.

[13]James P. Shaver, *A Study of Teaching Style: The Investigation Through Systematic Observation of the Ability of Experimental Teachers to Conform to Two Models of Teaching*, (Cambridge, Massachusetts: Harvard University Archives, 1961).

carried out to investigate relationships between teachers' ability to teach using different methods and certain teacher factors.

Fischler (1967) outlined an evaluative procedure he proposes to use in his attempt to discover whether a four week in-service training program will produce a change in the teachers' classroom behavior.[14] Although no results of the evaluation are presently available to this investigator, the proposed methods of evaluation hold promise of an objective approach to the problem of evaluating long term effects of a methods in-service program.

Worthen (1965) reported a study designed to investigate whether teachers were able to shift their teaching procedures to fit two specifically different teaching models.[15] By utilizing observers rating teacher clssroom behavior patterns and pupil inventories of teacher behaviors, he found that the eight participating teachers were able to shift their teaching patterns significantly during the experimental period.

---

[14] Abraham S. Fischler, "Change in Classroom Behavior Resulting from an In-Service Program Utilizing Television," School Science and Mathematics, Vol. LXVII, (April, 1967), pp. 321-324.

[15] Blaine R. Worthen, "Discovery VS Expository Sequencing in Six Weeks of Classroom Instruction in Mathematics," in Sequence Characteristics of Text Materials and Transfer of Learning, Gabriel M. Della-Piana, Garth M. Eldrige, and Blaine R. Worthen (Authors), (Bureau of Educational Research, University of Utah, Salt Lake City, 1965).

Cooksey (1965) investigated changes in student achievement in English as a result of teachers' participation in an in-service program.[16] As a secondary part of his study, he observed teachers in the classroom in an attempt to identify the ways in which they changed their teaching practices. Significant changes were indicated in the teaching practices for all ten teachers when considered as a combined group.

Dossett (1964) evaluated the effectiveness of workshops conducted in Missouri during 1963 and 1964 in terms of improvement of teachers' basic mathematical understanding, changes in teachers attitudes toward mathematics, and improvements in classroom practices in the teaching of arithmetic.[17] A random sample of two students from each of eight classes taught by teachers participating in the workshops were questioned to obtain data on changes in teacher practices. The result of questioning the students indicated that the teachers allowed more discussion, and involved pupils in more non-textbook activities as a result of the workshop in-service program.

---

[16] Henry B. Cooksey, "Evaluation of an In-Service Education Program for English Teachers," Unpublished Doctoral Dissertation, (University of Texas, 1965).

[17] Mildred J. Dossett, "An Analysis of the Effectiveness of the Workshop as an In-Service Means for Improving Mathematical Understandings of Elementary School Teachers," Unpublished Doctoral Dissertation, (Michigan State University, 1964).

Literature reviewed in this section indicated that the need for methods in-service education programs has been recognized. However, the main emphasis of in-service programs per se has been on upgrading teachers' competence in new mathematical content and structure rather than on enlarging the teachers' repertoire of teaching methods. Few studies were found by this investigator which were designed specifically to test the effectiveness of methods in-service programs. The results of the studies reviewed are not conducive to comparison or generalization.

II. REVIEW OF THE LITERATURE ON THE ANALYSIS AND EVALUATION OF CLASSROOM INTERACTION FOR METHODS OF INSTRUCTION

That research on teaching methods may be carried out in its own right was ably stated by Smith (1960) in the following words:

> That learning does not necessarily issue from teaching, that teaching is one thing and learning is quite another, is significant for pedagogical research. It enables us to analyze the concept of teaching without becoming entangled in the web of arguments about the process and conditions of learning; ...[18]

Wallen and Travers (1963) discuss three approaches that may be used in identifying methods of teaching. One technique is observing

---

[18] B. Othanel Smith, "A Concept of Teaching," Teachers College Record, Vol. LXI, (February, 1960), p. 233.

the teachers directly. Another is by asking teachers, through an interview or questionnaire technique, to describe how they behave in the classroom. A third approach is to have the pupil's record, through an inventory, what happens in a particular teacher's classroom.[19] Flanders (1965), however, in discussing teacher influence in the classroom, indicated that direct observation of teacher behaviors and use of pupil inventories are two of the more useful ways of assessing teacher behaviors.[20]

In this section, literature was reviewed on the three approaches that may be used to identify methods of teaching as suggested by Wallen and Travers. The investigator has summarized the procedures and findings of some recent research that in his opinion was based on clearly defined experimental procedures, drew appropriate conclusions, and was considered particularly applicable to this investigation.

Research on Observer Rating Scales

Observer rating scales, according to Flanders, have been based on one of two models. The first model makes use of logical

---

[19]Norman E. Wallen and Robert M. W. Travers, "Analysis and Investigation of Teaching Methods," Handbook of Research of Teaching, Nathaniel L. Gage, editor (Chicago: Rand McNally and Company, 1963), p. 468.

[20]Ned A. Flanders, "Teacher Influence, Pupil Attitudes, and Achievement," Studying Teaching, James Raths et al, editors (Englewood Cliffs: Prentice-Hall, Inc., 1967), p. 44.

steps in problem-solving. It draws heavily from what is known of inductive and deductive reasoning, scientific method, procedures for defining terms, and principles from the field of semantics. The second model makes use of the intellectual skills of the first, but is based more on a set of social skills used by teachers to control and manage class activities. It is based on the psychology of superior-subordinate relationships adopted to fit classroom activities.[21]

A variety of observer rating scales and techniques designed to enumerate classroom verbal interactions have been developed. Flanders (1960) reported a scale developed on the basis of his second model utilizing a ten by ten matrix for the analysis of direct and indirect teacher influence.[22] Through this analysis he has developed some insights regarding the sequential and predictable character of teacher-learner interactions.[23] Amidon's (1963) Verbal Interaction Category System is an improvement on the Flanders scale by differentiating or classifying the teacher and student discourse on a seven-

---

[21]Flanders, op. cit., p. 43.

[22]Ned A. Flanders, "Analyzing Teacher Behavior," Educational Leadership, Vol. XIV, (December, 1961), pp. 173-180.

[23]John Withall, "Conceptual Frameworks for the Analysis of Classroom Social Interactions," Journal of Experimental Education, Vol. XXX, (June, 1962), p. 307.

teen category scale.[24] Rather than get involved with value judgments on the merits of direct and indirect influence, a feature of the Flanders analysis, Amidon looks at the categories from the point of view of initiation and response.

Smith (1960) and associates designed a system in which classroom discourse was taped and analyzed into pedagogically neutral units called "episodes."[25] Each of the episodes was recorded under a particular category of the Opening Phase, Continuing Phase, or Closing Phase of the interaction. In this way, the complete verbal interaction of the classroom was categorized into phase, category and sub-category. A measure of the pattern or style of verbal communication behaviors used by the teacher in his instruction was obtained from the pattern recorded. This pattern or style was then analyzed and evaluated in terms of the soundness and logic of the thought processes.

Mitzel and Medley (1958) developed the Observation Schedule and Record technique (OScAR) which was based on the Withall Climate Index and on the system used by Cornell, Lindvall and Saupe.[26] The OScAR technique attempted to categorize teacher and pupil verbal behaviors within the emotional climate and social organization of the classroom.

---

[24] Edmund Amidon and Elizabeth Hunter, "Verbal Interaction in the Classroom - The Verbal Interaction Category System," Unpublished manuscript, (Temple University - Hunter College, 1963).

[25] B. Othanel Smith, "A Concept of Teaching," Teachers College Record, Vol. LXI, (February, 1960), pp. 229-241.

[26] Donald M. Medley, "Experiences with the OScAR Technique," Journal of Teacher Education, Vol. XIV, (September, 1963), pp. 267-272.

The task of the observer was limited to obtaining a record of the behaviors observed without attempting to evaluate it. This record was then analyzed and evaluated by a team of analysts in terms of teacher "effectiveness." Schueler, Gold, and Mitzel (1962) have used the OScAR in their teacher training programs with certain adaptations.[27] Ratings on the teacher's appearance, clarity of speech, and style of presentation were added to OScAR. Television and kinescope recorders operated by remote control were used to obtain complete records of the classroom activities in an attempt to overcome the effect of the presence of outside observers in the classroom.

Hughes and her associates (1959) developed a set of categories similar to those of Withall's. The observers, however, did not restrict themselves to recording verbal behaviors alone. Use was made of narrative anecdotal descriptions to obtain behavioral data that was greater in scope that would be possible with a categorized schedule alone.[28] Through the analysis of data thus obtained, Hughes attempted to formulate a model pattern for teacher behaviors. However, when applied to the thirty-five teachers of the major study group, none of

---

[27]Herbert Schueler, Milton J. Gold, and Harold E. Mitzel, The Use of Television for Improving Teacher Training and for Imprving Measures of Student-Teaching Performance, Phase I. (New York: Hunter College, 1962).

[28]Marie Hughes, Development of the Means for the Assessment of the Quality of Teaching in Elementary Schools, (Salt Lake City: University of Utah Press, 1959).

the seven teacher functions discriminated between the twenty-five "judged-good" teachers and the ten "representative" teachers.[29]

Bales and Gerbrands (1948) designed an ingeneous apparatus to move recording tape at a predetermined rate so that the record of observations was kept in its original temporal sequence.[30] The aim of the schedule developed by Bales was to obtain a series of indices which may be significant in representing small group dynamics. The Interaction Recorder has been used successfully by Perkins (1964) in his study of classroom interactions in underachieving classes. Perkins developed his own schedule of Teacher Categories and Student Categories for the purposes of his study.[31]

One of the few observation schedules designed to study interaction within a mathematics classroom was developed by Wright (1959). The categories, or frames of reference of the schedule (Ability to Think, Appreciation of Mathematics, Attitude of Curiosity and

---

[29] Donald M. Medley and Harold E. Mitzel, "Measuring Classroom Behavior by Systematic Observation," Nathaniel L. Gage, editor, Handbook of Research on Teaching, (Chicago: Rand McNally and Company, 1963), p. 271.

[30] Robert F. Bales and Henry Gerbrands, "The 'Interaction Recorder', An Apparatus and Check List for Sequential Content Analysis of Social Interaction," Human Relations, Vol. 1, (November, 1948), pp. 456-463.

[31] Hugh V. Perkins, "A Procedure for Assessing the Classroom Behavior of Students and Teachers," American Educational Research Journal, Vol. 1, (November, 1964), pp. 249-260.

and Initiative) related to the objectives of mathematics teaching.[32] An imaginative approach to recording observations was employed by Wright in that lessons were observed for a fifteen second period. The observed behaviors were then categorized during the next fifteen second period. This procedure was followed through the total length of the lesson observed. The data collected through the use of this schedule enabled Wright to analyze classroom discussions in terms of the degree of rigor of mathematics taught as well as the degree of student participation in the lessons observed.

Shaver (1961) in the review of literature for his study of teaching style found that none of the studies he reviewed used systematic observation as a method of operationally defining the teaching procedures or methods used.[33] He developed an observational system in which teacher and student behaviors were classified in such a way as to give a measure of the extent to which predefined experimental methods were in use. The extent to which an experimental method was used was determined on the basis of the deviation of the observed score from the score of a theoretically ideal method.

---

[32]Muriel J. Wright, "Development of an Instrument for Studying Verbal Behaviors in a Secondary School Mathematics Classroom," Journal of Experimental Education, Vol. XXVIII, (December, 1959), pp. 103-121.

[33]Shaver, op. cit., p. 8.

Worthen (1965) used a somewhat similar approach in the observation of teachers in his study of discovery versus expository sequencing of mathematics instruction.[34] The Rating Scale was divided into Discovery-Expository Items and General Items. Five teacher behavior dimensions were rated under the first set of items and six under the second set. Each dimension was rated on a six-point Discovery-Expository continuum. A measure of the extent to which the teachers were able to adjust their teaching methods to fit the predetermined models was indicated by the size of the positive increment of the experimental period scores over the pre-experimental scores for the Discovery Mode, and the size of the negative increment of the two scores for the Expository Mode.

Research on Pupil Inventories

Several attempts have been made to enlist pupils in the task of determining the nature of classroom activities. Measures of pupil perceptions have been generally obtained through the use of inventories or questionnaires. Cogan (1958) developed such a questionnaire called Pupil Survey which he used as a major source of data concerning the work of the pupils and the behavior of their

---

[34] Worthen, *op. cit.*, p. 509.

teachers.[35] Cogan found that although the pupils' rating of their teacher's behavior were highly stable, they were not consistently related to the principals' rating of the teachers' behaviors. There is some question as to whether the frames of reference the pupils used were the same as those used by the principals. The evidence does appear to support Traver's contention that pupils' perception may well be different from that of an adult observer.[36]

The Texas Classroom Interview Question Schedule was developed in a study conducted by DeVault, Houston, and Boyd (1962) after the manner used by Shannon and Wishard to elicit responses from six pupils chosen at random from one class of each of forty-five teachers. The pupil interview technique was used in an attempt to evaluate the classroom practices of the teachers after the teachers participated in an in-service program conducted by the investigators.[37] Pupil

---

[35] Morris L. Cogan, "The Behavior of Teachers and the Productive Behavior of Their Pupils: 1. 'Perception' Analysis," Journal of Experimental Education, Vol. XXVII, (December, 1958), pp. 89-105.

[36] Robert M. W. Travers, An Introduction to Educational Research, Second edition, (New York: The Macmillan Company, 1967), p. 277.

[37] M. Vere DeVault, Robert Houston, and Claude Boyd, Television and Consultant Services as Methods of In-Service Education for Elementary School Teachers of Mathematics, (Bureau of Laboratory School Publications No. 15. Austin, Texas: University of Texas, 1962).

responses were elicited on the following components: (1) materials, (2) classroom activities, (3) organization, (4) teaching aids, (5) evaluation. The questions were phrased in such a way as to avoid any biasness of the pupils toward their teacher. The results of the oral questionnaire technique provided data sufficiently reliable for the researchers to conclude that the in-service program utilizing television and lectures with face-to-face discussions was more effective than the other three methods used. Dossett (1964) successfully used the Texas Classroom Interview Question Schedule and an accompanying Rating Scale to elicit responses from a random sample of students as the major source of data on teacher classroom practices.[38] From the analysis of the data so obtained, she concluded that the teachers as a result of the in-service programs conducted in the State of Missouri, allowed more discussion, involved pupils in more non-textbook activities, andused more non-text materials.

Worthen (1965) used a pupil inventory consisting of twenty-five items which were designed to give measures of the degree to which the teachers followed the expository or discovery models in their teaching.[39] As this technique was used to corroborate the

---

[38] Dossett, op. cit., p. 107.

[39] Worthen, op. cit., p. 545.

findings of the observers, the questions were based on the same teacher behavior dimensions used in the construction of his Observer Rating Scale. A negative response typified "low discovery" while a positive response typified "high discovery." The results of this inventory did corroborate the findings of the raters using the Observer Rating Scale.

Teacher Opinionnaires

Although Wallen and Travers have stated that one way of obtaining evidence on the nature of classroom behaviors is to ask the teachers to describe their behaviors in the classroom[40], few researchers have used this technique. Roscoe used a teacher opinionnaire to obtain data on the effect of the Alabama Polytechnic Institute workshops on the participating teachers' teaching methods.[41] Hilgert (1968) elicited teacher reactions to the summer workshops conducted by the Washington University in Economic Education.[42] To evaluate the impact of these workshops upon the teachers' teaching, Hilgert con-

---

[40] Supra, p. 22.

[41] Roscoe, op. cit.

[42] Raymond L. Hilgert, "Teacher Reaction to Summer Workshops," School and Community, Vol. 54, (January, 1968), pp. 14-15.

ducted a mail questionnaire survey of seventy teachers. Responses to the semi-structured questions indicated that the three-week workshop had an impact on some of the teachers' methods from the point of view of materials used, classroom projects, guest speakers, and field trips although the workshop programs did not emphasize methods of teaching.

Other studies (Jirik, 1964)[43] used this technique to elicit teacher opinions as to the extent certain features of in-service programs were being used in classrooms and the possible effect of these features on pupil achievement.

Wallen and Travers point out certain obvious pitfalls of this technique.

> One is that the teachers, like other persons, probably have only limited insight into what they do and hence will record their concept of how they behave in the classroom rather than what they actually do. Another is that teachers have difficulty in recalling just what they did in the classroom or how much time was devoted to this and that activity.[44]

In another source Travers stated that:

> Opinions given on the spur of the moment are likely to be prompted by incidents that would have little weight after the subject had been given careful consideration.[45]

---

[43] Edward F. Jirik, "A Statistical Analysis of 'Level of Achievement' as Related to Certain Supervisory Services and In-Service Participation," Unpublished Doctoral Dissertation, (Ohio University, 1964).

[44] Wallen and Travers, op. cit., p. 468.

[45] Robert M. W. Travers, An Introduction to Educational Research, Second edition, (New York: The Macmillan Company, 1967), p. 288.

Thus, the time dilemma faces researchers attempting to use this technique for gathering data on classroom interaction. If the opinionnaire is administered too soon, insignificant incidents may colour responses; if the opinionnnaire is administered too long after instruction, significant incidents may be forgotten.

The review of the literature in this section revealed that assessment of the teaching process was most frequently carried out with the intention of identifying certain factors conducive to good or effective teaching rather than identifying methods of instruction. Although the development of observation schedules was the more popular concern of "professional" investigators, pupil inventories and teacher opinionnaires were used to good advantage by researchers who deemed those techniques adequate.

III. SUMMARY

Those aspects of the literature reviewed deemed pertinent to the present study are summarized as follows:

1. In recent years there has been a tremendous effort exerted on upgrading the teachers through in-service programs.
2. There is some evidence in the literature indicating attempts to objectively evaluate the effectiveness of these in-service programs.

3. In-service training programs designed to acquaint teachers with new methods of teaching and to enlarge their repertoire have not been as common as in-service training programs designed to improve the teachers' competence in subject matter content.
4. Assessment of the teaching process has been generally based on direct observations of classroom interactions or indirect evaluation of responses obtained from pupils and teachers.
5. Generally, classroom observations of specific acts of teachers and students have been made to assess teaching effectiveness, group interactions, or cognitive operations.
6. Procedures for recording what goes on in a classroom have been formulated on the basis of anecdotal or descriptive statements or on a category, classification system.
7. Records from direct observations appear to be more detailed and precise than those obtained from opinionnaires or questionnaires.
8. The constructs of observation schedules designed to obtain data on classroom interactions per se were not directly related to the constructs necessary to an operational definition of the two teaching methods used in this study.

9. Practically no research evidence is available on the relation between certain teacher variables and the extent of flexibility of the teachers to adopt different teaching methods.

CHAPTER III

EXPERIMENTAL DESIGN AND PROCEDURES

The main purpose of this study was to provide research evidence that could be used as a basis for evaluating the relative potential of an in-service program for effecting change in the teaching methods of certain teachers of Mathematics 20. A secondary purpose was to investigate and develop instruments whereby the extent to which teachers changed their teaching methods could be measured objectively.

I. THE NATURE OF THE SUBJECTS

The subjects for the experiment were selected from those teachers who were teaching at least two classes of Mathematics 20 in the Edmonton Public School System during the 1967-1968 academic school year, and who volunteered to take part in the experiment. An additional two teachers were used in the pilot study. Ideally, it would have been highly desirable to have a random sample of Mathematics 20 teachers. However, since a determining factor was that the teachers were required to be teaching at least two classes of Mathematics 20, while a practical limitation was the desirability of having teachers at approximately the same point in the course at the end of January, such a sample was not possible. Table I on the following page summarizes the selected teacher characteristics of the subjects selected.

TABLE I

SUMMARY OF SELECTED CHARACTERISTICS OF THE PARTICIPATING TEACHERS IN THE MAIN STUDY

Teacher Characteristics	1	2	3	4	5	6	7
Age	22	29	28	29	27	47	34
Sex	F	F	M	M	M	F	M
Years of Teaching Experience	1	8	5	5	4	18	14
Years of H. S. Math. Exp.	1	7	5	5	2	18	7
No. of Times Math. 20 Taught	2	4	3	6	3	3	5
No. of Univ. Math. Courses	5	6	3	4	7	8	9
No. of Degrees	1	1	2	2	2	2	3
Recency of Tr.	67	67	67	62	65	67	66
M.T.A.I. Rank	56	63	46	8	20	37	67[#]

[#]Percentile ranks obtained by the teachers on the Minnesota Teachers Attitude Inventory as determined by the norms for Academic Secondary Teachers with appropriate years of training.

Teachers were required to complete the Teacher Characteristics Questionnaire[1] and the Minnesota Teacher Attitude Inventory as a source of data for Table I. The number of times the teachers had taught Mathematics 20 was taken to be the number of classes of the present Mathematics 20 authorization that the teachers had taught.[2] All teachers have indicated that they had taken at least two University mathematics courses since 1962, which indicates that the recency of their mathematics training is quite comparable. In an attempt to place the teachers on an Authoritarian-Permissive continuum, the teachers were asked to complete the Minnesota Teacher Attitude Inventory.[3] This Inventory has been found by Kingston and Newsome (1960) to compare significantly with the WSF Scale developed by Webster, Sanford, and Freedman and with the Inventory of Classroom Administrative Philosophy (ICAP) developed by Kingston.[4] The items

---

[1]See Appendix A.1.

[2]W. B. MacLean et al, Secondary School Mathematics Grade Eleven, Alberta edition, (Vancouver: Copp-Clark Publishing Co. Ltd., 1966).

[3]Walter W. Cook, Carrol H. Leeds, and Robert Collis, Minnesota Teacher Attitude Inventory, New York: Psychological Corp., 1951.

[4]Albert J. Kingston and George L. Newsome, "The Relationship of Two Measures of Authoritarianism to the Minnesota Teacher Attitude Inventory," Journal of Psychology, Vol. XVIX, (February, 1960), pp. 333-338.

of the M.T.A.I. appear to measure attitudes which reflect on the interpersonal-relationships between pupil and teacher and may be considered indicative of authoritarian patterns of behavior or belief.[5] The items of the WSF and ICAP scales are believed to be more personality centered and were designed to determine the degree to which teacher's practices and beliefs reflect the concept of democracy. All three tend to appraise some aspects of the authoritarian "syndrone."[6] The percentile ranks are so arranged that the lower the percentile rank, the more authoritarian is the indicated personality.

An inspection of Table I shows that the teachers participating in the main experiment are relatively well trained in recent mathematical ideas. There is an appreciable range in age, years of experience in teaching, and in teaching high school mathematics represented in the group. The range represented by the percentile ranks for the authoritarian-permissive characteristic as indicated by the M.T.A.I. appears to be quite substantial.

II. EXPERIMENTAL DESIGN AND CONTROLS

Seven teachers, each having at least two Mathematics 20

---

[5] Ibid.

[6] Ibid.

classes served as experimental subjects. Each teacher was required to teach one class using the Mathematizing Mode, while the other class was taught by the Expository Mode. This was done in order to control for the dimensions of teacher personality and other teacher characteristics.

The experimental period was approximately seven weeks in length, beginning the first week of February and ending at the end of March. Each class was taught Chapters 9 and 10 on the quadratic function of the present Mathematics 20 course.

The major non-experimental variables controlled in this study are presented below.

1. The research design balanced, as far as possible, the achievement level of the students assigned to each treatment. An attempt was made to match the classes assigned to each treatment on the basis of the teachers' rating of the classes' previous achievements.
2. Each teacher devoted approximately seven weeks in covering the two units on quadratics in each treatment.
3. Although the type of verbal behavior varied to fit the two teaching models, the kind of teacher verbalization in the Mathematizing Treatment was controlled to the extent that the teachers were able to operationally interpret the classroom activities outlined in the hand-outs prepared for them.[7]
4. In order to obviate the criticism that the observations of teacher classroom behaviors were biased by the investigator being the only observer, student rating of teacher behaviors was solicited as another source of data to augment that obtained through the investigator's observations.

---

[7]Ross Johnston, "The Mathematizing Mode," (A Working Paper, University of Alberta, Edmonton, 1968).

5. The research design and all the various procedures attempted to negate any differential "Hawthorne Effect" between the two methods by designating each as an experimental mode.

III. THE IN-SERVICE EDUCATION PROGRAM

A five session methods in-service education program was conducted which was attended by all participating teachers and this investigator. The training class met for a minimum of two hours every two weeks during the months of February and March, 1968. Training was given in two areas. First, the theoretical framework of the Mathematizing Mode[8] was explained to all teachers in a face-to-face lecture-discussion session. Second, instructional materials, developed and prepared by Johnston were handed out to the teachers. These materials, in the form of "classroom activities," were based on several mathematical concepts from the two units on the quadratic function not yet covered by the classes. The materials were intended to give the teachers a set of instructional procedures to be followed in teaching by the mathematizing method. The classroom activities were explained and discussed during the training sessions. Although only one part of one session was devoted exclusively to a discussion of the Expository Mode, comparisons and contrasts between the two methods were drawn constantly during all sessions. During the initial experimental period, the teachers were

---

[8] Ibid.

visited in the classrooms by the directors of the training program in an effort to provide the teachers with further help with the mathematizing mode.

Teachers were requested to complete the investigator prepared Post-Meeting In-Service Evaluation Sheet[9] after each session. This Sheet was based on one used by O'Rourke (1957) in her study of 261 administrators of 43 states participating in 195 workshops.[10] The Sheet was designed in an attempt to provide immediate feedback on the degree to which the in-service session met with the expectations of the teachers. Suggestions on the in-service program were solicited and an attempt was made to incorporate them into the subsequent sessions.

IV. COLLECTION OF DATA ON TEACHERS' ADHERENCE TO INSTRUCTIONAL MODELS

Teacher classroom behaviors were rated in two ways. First, this investigator observed each of the teachers in action in each class on two occasions before the experimental period, during the last two weeks of the experimental period, and approximately six

---

[9] See Appendix A.2.

[10] Mary A. O'Rourke and William H. Burton, Workshop for Teachers (New York: Appleton - Century - Crofts, Inc., 1957), p. 74.

weeks after the experimental as indicated in the following schedule.

	Observations by Investigator	Student Inventory
Pre-Experiment (Jan. 16-31)	Two observations of each class in each treatment	One completion of Student Inventory
During Experiment (March 8-25)	Two observations of each class in each treatment	One completion of Student Inventory
Post-Experiment (May 22-June 7)	Two observations of each class in each treatment	One completion of Student Inventory

The Observer Rating Scale of Teacher Behavior described in Chapter IV was used during each observation to obtain data on the teacher's classroom behaviors. The students were asked to complete the Student Inventory of Teacher Behavior described in Chapter IV once during each of the above periods. This procedure provided the investigator with two sets of data on teacher behaviors before the experiment, during the experiment, and after the experiment. Ideally, it would have been highly desirable to have multiple ratings on the same lessons based on live observations of teachers in action or on video tapes of a random sample of the lessons taught in each treatment class. However, time and resources did not allow the training of a team of observers nor the video taping of teachers' lessons.

Observational Procedures

Travers (1958) pointed out one problem with observation in

the classroom.

> We can no longer accept the notion based on wishful thinking, that the introduction of an observer into the classroom does not affect events therein, for clearly it does.[11]

In an attempt to minimize the effect on the classroom interaction of the investigator's presence, two procedures were followed. First, the investigator made a point of getting to know the teachers prior to the period covered by this study. It was pointed out that the ratings would not be analyzed from the point of view of teacher effectiveness. Second, the investigator made a point of coming to the classroom before the class assembled to chat with the teacher so that the students would see that a friendly professional relationship existed. It was felt necessary, through this action, to create the impression that the teacher was not being "inspected." Upon being introduced to the class, the investigator informed the students that he was interested in observing how certain things were being done in class. From the back of the room, the investigator made observations of teacher and student activities and rated his interpretations of these behaviors on the Observer Rating Scale.

Administration of the Student Inventory

Students were asked to complete the Student Inventory of Teacher Behaviors at the beginning of the second lesson being

---

[11]Robert M. W. Travers, An Introduction to Educational Research, First edition (New York: The Macmillan Company, 1958), p. 358.

observed in each treatment class before, during, and after the experimental period.[12] Students indicated their responses to the 30 items of the Student Inventory on an Optical Scoring sheet. Since the investigator was interested only in the class consensus rating[13] on each item, students were required to identify their response sheets by class only. Instructions to the students included the request that they refrain from discussing the items of the Inventory until after the response sheets were collected so that responses would reflect individual student's impressions. Each administration of the Inventory required approximately 20 minutes of the class time.

V. STATISTICAL PROCEDURES

For the purposes of this study, data was collected on two teacher dimensions. Data on the first teacher dimension, selected teacher characteristics, was obtained through the Teacher Characteristics Questionnaire and the Minnesota Teacher Attitude Inventory. Data on the second teacher dimension, teachers' adherence to the theoretical treatment models, was gathered through the Observer Rating Scale and the Student Inventory of Teacher Behaviors.

The null hypotheses tested in this investigation were:

1. There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes before the in-service treatment.

---

[12] See Schedule, p. 43.

[13] *Infra*, p. 60.

2. There is no significant difference between teachers methods ranks as indicated by the Teachers Category Index scores before the in-service treatment.
3. There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes during the experiment.
4. There is no significant difference between teachers methods ranks as indicated by the Teacher Category Index scores during the experiment.
5. During the experiment, there is no significant difference between ranks of teachers when ranks are determined by:
   a) age and flexibility of methods
   b) years of teaching experience and flexibility of methods
   c) number of times they had taught Mathematics 20 and flexibility of methods
   d) number of university mathematics courses taken and flexibility of methods
   e) number of university degrees held and flexibility of methods
   f) percentile scores obtained on the Minnesota Teacher Attitude Inventory and flexibility of methods.
6. There is no significant difference in the ranks of flexibility of methods between sexes during the experiment.
7. There is no significant difference in a teacher's methods for classes before the experiment and after the experiment.
8. There is no significant difference between teachers' methods for classes after the experiment.

In testing the above null hypotheses the data collected on both teacher dimensions were used as a basis for ranking the seven teacher subjects. Suitable nonparametric tests were considered appropriate in testing the null hypotheses for several reasons. First, since the seven subjects in the main study, were not randomly drawn from the population of Mathematics 20 teachers, one could not

assume the criterion of normality for the experimental group.[14] Second, the investigator had no basis on which to assume the normality of variance of the teachers' adherence to the mathematics models. Third, the scores obtained on the two instruments appear to have relative meaning only and hence, may be assumed to have the strength of ordinal measures only. According to Ferguson (1966) and Siegel (1956) under these conditions nonparametric statistical tests may be appropriately used.[15]

Tests of null hypotheses 1 and 2 were carried out on ranks of teachers according to the scores they received from the various sources described above. Tests of null hypotheses 3 to 6 were carried out on the ranks of teachers based on the differences between the scores obtained during the treatment period and before the treatment period. Tests of hypotheses 7 and 8 were carried out on the ranks of teachers based on the differences between scores received after the treatment period and before the treatment period. The latter two procedures were used in an attempt to adjust for initial differences within the teachers' methods and between the

---

[14] B. J. Winer, Statistical Principles in Experimental Design, (New York: McGraw-Hill Book Company, 1962), p. 5.

[15] George A. Ferguson, Statistical Analysis in Psychology and Education, (New York: McGraw-Hill Book Company, 1966), p. 355 and Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, (New York: McGraw-Hill Book Company, 1956), p. vii.

teachers' methods. The actual computations involved were carried out by utilizing appropriate Fortran and APL/360 computer programs written by the investigator for this purpose with the help of the Division of Educational Research Services of the Faculty of Education.

The present chapter described the nature of the various components of the experiment, of the experimental design and controls, and of the general statistical procedures followed in analyzing the data. Chapter V presents the specific statistical tests used and the consequent results of the analysis of the data.

CHAPTER IV

PREPARATION OF THE INSTRUMENTS

This chapter contains a statement of the rationale underlying the instruments developed for the purposes of this study, a detailed description of the instruments, and a description of the pilot study conducted to test the instruments.

I. MEASURES OF TEACHER CLASSROOM BEHAVIORS

Rationale

One of the problems of this study was to investigate and develop instruments and procedures through which the effectiveness of a methods in-service program for changing teachers' methods of instruction could be measured objectively. The constructs of the observation schedules used in the studies reviewed in the literature did not appear to be sufficiently related to those needed to operationally differentiate between the two experimental treatment models. It was necessary for this investigator to develop two instruments in an attempt to assess the extent to which the teachers were able to follow the expository or mathematizing models in their teaching.

The two instruments developed were based on the following assumptions:

1. One of the better means by which data may be obtained on teaching methods is through observations of what goes on in the classroom.
2. High school students are keen observers of what the teacher does in the classroom and how he does it.
3. Data obtained from a Student Inventory of Teacher Behaviors which did not require students' judgment of teacher's competence or effectiveness is comparable to that obtained by an outside observer rating teachers on the same basis.

The Observer Rating Scale of Teacher Behavior was used by the investigator while observing each of the teachers in action in each of the two classes. The Student Inventory of Teacher Behavior was in a form of a questionnaire and administered to all the students of the two classes selected for each teacher. Each of the instruments were adapted from similar instruments used by Worthen (1965) in his investigation.[1]

Six dimensions of teaching behaviors were used as a basis for each of the two instruments. These were selected from among the many behaviors that appear to differentiate the mathematizing method from the expository method of teaching. Each of the six dimensions are observable to some degree in most lessons taught by either method. They therefore form the basis of an operational definition of the two methods under study. Each of the dimensions became a distinct category on the Observer Rating Scale. Five

---

[1]Worthen, op. cit., Part I, p. 5.22.

correlated items were rated under each category to form a composite score for that category. This procedure, according to Medley,[2] provides for a more reliable and interpretable score.

The thirty items of each instrument were such that they did not require judgments of the teachers' competence or effectiveness in the use of either method.

Ideally, the two teaching methods, expository and mathematizing, were considered to be extremes on a bi-polar continuum. Ratings on this continuum were facilitated through the use of a five-point scale. The five-point scale has been found (Bryan and DuBois, 1961) to be superior to a four-point scale for obtaining student reactions to teachers.[3] Scores obtained from each instrument represented a comparable record of general impressions of what went on in the classes rather than a quantified enumeration of specific units of teacher behaviors.

Observer Rating Scale of Teacher Behavior

This rating scale[4] was designed to be used by the investigator

---

[2]Donald M. Medley, "Experiences with the OScAR Technique," Journal of Teacher Education, Vol. XIV, (September, 1963), p. 268.

[3]Roy C. Bryan and Karl DuBois, "Comparison of Two Rating Scales Used to Obtain Student Reactions to Chemistry Teachers," School Science and Mathematics, Vol. LXI, (March, 1961), pp. 175-180.

[4]See Appendix B.1.

while observing teachers in the classroom. The extent to which teachers were observed to follow each of the treatment models was rated on the basis of the following operational descriptions.

The Mathematizing Model of teaching is described briefly under each category.

1. Teacher Omniscience: In the mathematizing method of teaching, the teacher consciously avoids any behavior which would establish him as the primary source or authority of mathematical knowledge. He acts in the capacity of a resource person who indicates the direction in which the enquiry might take to arrive at a possible solution rather than show the students how to solve the problem. In checking work or summarizing a problem-solving procedure, the teacher using the mathematizing method relies on student responses or suggestions rather than on the most "direct way" of arriving at the solution. Feedback on the correctness of the solutions comes from the mathematics rather than from the teacher. The teacher using the mathematizing method does not deliberately have responses corrected as the lesson progresses.

2. Introduction of Generalization: Generalizations are end products of student enquiry into various facets of a problem situation in the mathematizing approach to teaching. In this method, examples are used for the purpose of "testing out" hypotheses formulated by the students. The teacher encourages the students to hypothesize at

procedures to solutions. After the problem situation has been explored and most of the students have become aware of the mathematical ideas relevant to the solution, the teacher leads the class to summarize the concepts of the problem situation and formulate generalizations.

3. Control of Pupil Interaction: In the mathematizing method, pupil interaction relative to the generating of hypotheses is encouraged. That interaction, however, which may result in preventing individual students from actively inquiring into the problem situation in the hope of "discovering" the principle involved is discouraged until after the verbalization of the rule is drawn from the class. Through appropriate suggestions and questions the teacher mediates classroom interaction to explore the problem situation, formulate and test the hypotheses, and to summarize the principles involved.

4. Method of Answering Questions: When a question is asked of the teacher he assumes the role of an adviser. That is, through re-wording or re-grouping of examples, he leads the students to consider those aspects of the computational sequence which will enable them to recognize the mathematical principle that is necessary to the problem's solution. In referring to a student's computational sequence, the teacher asks questions designed to reinforce the student's understanding of the principle that has been overlooked or

incorrectly applied. The teacher does not show the student how to solve the problem.

5. Use of Student Responses: Student responses are accepted without evaluation by the teacher. The class evaluated the responses in relation to the case at hand. After due consideration by the class, the teacher takes advantage of the responses to guide classroom interaction toward the objectives of the lesson. Productive student responses in the mathematizing method are those which suggest possible hints to the solution of the problem. These are the responses which are encouraged and used in the classroom discourse.

6. Method of Eliminating False Concepts: In the mathematizing method, problems are assigned to the students without any indication as to which have special peculiarities or those problems to which a particular concept does not apply. If a student discovers that a particular rule or approach does not work, the teacher does not explain why the rule does not apply but allows the students to obtain feedback from mathematics to find the reason. Problems which deliberately lead the class to overgeneralize the rule are included in the assignments. Students are allowed to try out a variety of methods in solving problems so that they may have a basis on which to compare and select that approach they deem most appropriate.

<u>The Expository Model</u> of teaching as revealed by observable teacher behaviors on each of the categories is described in the following way.

1. Teacher Omniscience: In the expository method of teaching, the teacher makes a conscious effort to project to the students an attitude of confidence in, and authority with, the mathematics under discussion. Because students feel the teacher is competent in the mathematics they seek his help to solve problems for them. The teacher solves problems directly, using appropriate procedures. Solutions to problems are corrected with necessary explanations before proceeding.

2. Introduction of Generalization: The teacher introduces and identifies the specific topic to the class in the expository method. Rules or principles are presented and explained first by the teacher, after which they are illustrated by example problems. The teacher summarizes the application of the rules after each subtopic or set of examples. Generalization of the principles or rules concludes the lesson sequence.

3. Control of Pupil Interaction: The interaction of the class is directed toward the presentation of a "correct and proper" application of the rule in an appropriate problem-solving situation. The problem-solving procedure is broken into a number of "logical" steps. Student attention is directed at this series of steps. The teacher

questions students in such a manner that the students in responding will reveal the rules that apply to the direct solution of the problem. The students are free to ask any questions or offer any suggestions pertaining to the application of the rules or principles that may be helpful to the understanding of the generalization.

4. Methods of Answering Questions: When a question is asked of the teacher in the expository method, the teacher re-states the rule and gives further examples in an attempt to clarify it. If the student is still unable to understand the application of the rule, the teacher may then show the student how to solve the problem.

5. Use of Student Responses: In the expository method the teacher comments on or evaluates student responses without testing them in their proper context. Student responses are elicited as a source of feedback on whether the students are following the exposition of the lesson. On the basis of the responses the teacher may re-explain or re-illustrate the rules and their application.

6. Method of Eliminating False Concepts: In the expository method, the students are warned of the common errors made in applying the rules or principles. Problems with unusual peculiarities or hidden "tricks" are pointed out to the students. Specific suggestions are given students as to the way in which the problems are to be solved. Suitable problems are chosen to provide students with a variety of applications of the rule or principle discussed without over-generalizing it.

Scoring the Observer Rating Scale is accomplished in the following way. The extent to which each item was observed in the in the classroom was rated on a five-point weighted scale. Some of the items were worded from the mathematizing point of view while others were worded from the expository point of view in an attempt to avoid a mental set being established while using the scale. Rating weights were adjusted so that the teacher following the ideal expository method would be rated zero, while the teacher following the mathematizing method would be rated four on each item. The Teacher Category Index was obtained by summing the ratings of the five items under the category and multiplying this sum by five. This procedure rates the teacher's adherence to the expository or mathematizing methods for each category in multiples of five from zero to one hundred respectively. The Teacher's Method Index was obtained by finding the arithmetic mean of the Category Indexes.

To illustrate, suppose teacher X received the following Observer Rating scores for the third Teacher Category.

		YES				NO
C.	Control of Pupil Interaction					
1.	Students are encouraged to offer complete, correct solutions.	0	1	2	X	4
2.	Teacher questions students on direct items of the topic.	0	1	2	X	4
3.	Classroom interaction deviates from the topic.	4	X	2	1	0
4.	Students are encouraged to reveal rules that apply.	0	1	2	3	X
5.	Premature verbalization of the rule is discouraged.	4	X	2	1	0

Sum of the ratings on this category is 16.

This Teacher Category Index is 16 x 5 = 80.

Further, suppose the Teacher Category Index ratings for this lesson were as follows:

Category:	A	B	C	D	E	F
Index:	60	70	80	60	85	55

Sum of Indexes is 410.

The Teacher Method Index is 410 ÷ 6 = 63.3.

Student Inventory of Teaching Behaviors

This inventory[5] was filled out by all students of the classes involved in the experiment. Since the investigator was interested in obtaining the class impression of what went on in the classroom, students individually were not required to identify their response sheets. The thirty items of the inventory were designed to parallel the items of the Observer Rating Scale. However, the items were worded so that the students would not feel that they were rating teachers but rather that they were identifying the extent to which certain things happened in the classroom.

To aid the students in rating each item, the five-point scale was further defined as Almost Always, Often, Sometimes, Seldom, and Almost Never. Items 4, 7, 9, 10, 11, 12, 13, 14, 15, 20, 21, 22, 25, 27, and 29 were worded from the mathematizing point of view,

---

[5]See Appendix C.1.

while the other items were worded from the expository point of view in an attempt to avoid biasness toward any one method. The order of the wording of the five-point scale for each item was such that choises to the extreme left indicated the ideal expository method while those to the extreme right indicated the ideal mathematizing method of teaching.

Scoring the Student Inventory of Teacher Behaviors was carried out in the following manner. The rating of the teacher's behavior for the first Teacher Category Index (Teacher Omniscience) was obtained by finding the mean of the class consensus ratings of the first five items of the inventory. The class consensus ratings of the second five items were used to arrive at the second Teacher Category Index (Introduction of Generalization) and so on for each of the six Teacher Category Indexes.

The method used in arriving at the class consensus rating for each item was adapted from that used by Worthen.[6] The students' responses of each class were counted for each item and the percentages of the students answering the items each way were calculated. The order of the response choices for an item worded from the mathematizing point of view was reversed from that of the items worded from the expository point of view so that the weighting factors would be in the 0, 1, 2, 3, 4 order for all items. The

---

[6]Worthen, op. cit., Appendix A3.39.

weighted percentages were added and the sum was divided by 4. The resultant quotient became the class consensus rating for the item, with a possible range of zero for the ideal expository end of the continuum and one hundred for the ideal mathematizing end of the continuum.

To illustrate, suppose the item was number four.

4. When we ask the teacher how to solve a problem, he _____ gives us only hints that we may use.

   A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

Suppose the item was responded to by 34 students.

	Almost Never	Seldom	Sometimes	Often	Almost Always
Frequency	11	14	7	2	0
%	32.35	41.18	20.59	5.88	0
Wt.	0	1	2	3	4
Product	0	41.18	41.18	17.64	0
Sum		100.00			

Class consensus rating: $100.00 \div 4 = 25$.

II. PILOT STUDY TO TEST THE INSTRUMENTS

Development of the Initial Version of the Instruments

Copies of the First Draft of the two instruments, based on the rationale outlined in the preceding pages of this chapter, were circulated among five graduate students and two university professors of the Department of Secondary Education in Mathematics. Each of

these people were very familiar with the various aspects of the experiment. Critical appraisal of the items was invited, and in addition, each was asked to indicate whether in his opinion the items were worded from a high expository or high mathematizing point of view. On the basis of the comments received, many of the items were clarified by rephrasing, several inappropriate items were deleted, and a number of new items were added. The first draft of the two instruments so revised was then tested in a pilot study prior to the experimental period.

Pilot Study Procedure

Two classes taught by two different teachers not part of the main experiment were selected for the pilot study. One class was taught in accordance with the materials developed for the mathematizing method, while the other was considered to be taught by a method approximating the expository method. The investigator observed each class and rated the teacher's classroom behaviors on the Observer Rating Scale. On the following day, the Student Inventory was administered to all the students. The students were asked to indicate their responses on the basis of what they thought went on in class in the preceding two weeks. After a period of three weeks, the same procedure was followed.

Results of the Pilot Study

Tables II and III, which appear on the pages 63 and 64 present summaries of the ratings of the two teachers' classroom behaviors as obtained on the Observer Rating Scale and the Student Inventory respectively. Although these scores appear to differentiate between the mathematizing and the non-mathematizing teacher, there is some question whether the measurement of the Teacher Category variables are in an interval scale. This investigator has confidence that the scores are in at least an ordinal scale. Siegel suggests that when this situation exists nonparametric statistical tests should be applied to the data.[7]

Data in Tables II and III were subjected to nonparametric statistical tests in an attempt to answering the following questions:

Question 1. Did the two teachers use different methods as revealed by the scores when ranked separately for each category?

Since one expects some variation in rating from observation to observation,[8] the data was first sub-divided into the following pairs:

---

[7]Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, (Toronto: McGraw-Hill Book Company, Inc., 1956), p. vii.

[8]Joy P. Guilford, Fundamental Statistics in Psychology and Education, Fourth edition, (New York: McGraw-Hill Book Company, 1965), p. 450.

TABLE II

SCORES OBTAINED FROM THE OBSERVER RATING SCALE IN THE PILOT STUDY

Category	Mathematizing		Non-mathematizing	
	Obs. Rating #1	Obs. Rating #2	Obs. Rating #1	Obs. Rating #2
A	65	70	20	25
B	70	75	15	30
C	60	60	40	30
D	65	70	25	35
E	65	65	45	55
F	55	65	15	25
Method Index	63.3	67.5	26.7	33.3

TABLE III

SCORES OBTAINED FROM THE STUDENT INVENTORY
IN THE PILOT STUDY

	Mathematizing		Non-mathematizing	
Category	S. Inv. #1	S. Inv. #2	S. Inv. #1	S. Inv. #2
A	53	66	16	23
B	62	63	23	25
C	61	67	40	46
D	62	69	28	29
E	66	80	53	49
F	54	63	22	21
Method Index	59.7	68.0	30.3	35.5

1. The first Observer Rating Scores for the two teachers.
2. The second Observer Rating Scores for the two teachers.
3. The first Student Inventory Scores for the two teachers.
4. The second Student Inventory Scores for the two teachers.

The Wilcoxon Matched-Pairs Signed-Ranks Test[9] for differences in matched pairs of scores was carried out. This test gives more weight to a pair which shows a large difference between the two conditions than to one which shows a small difference. Its power efficiency is about 95.5 percent of the parametric t test.[10] The results of the Wilcoxon test on all four sets of pairs of scores led to rejection of the null hypothesis of no difference between the matched pairs of scores at the .05 level of significance since in each case T was equal to 0 with N = 6. Figures 1 to 4 on pages 66 and 67 show profiles of the matched pairs of scores tested. An examination of these profiles shows that no scores for one teacher overlap with those for the other teacher. This provides some evidence for the assumption that if the two teachers were in fact using two different methods then this difference would be reflected on the scores obtained from the two instruments.

Question 2. Are the scores obtained from the two instruments essentially the same for each teacher?

---

[9]Siegel, *op. cit.*, pp. 75-83.

[10]*Ibid.*

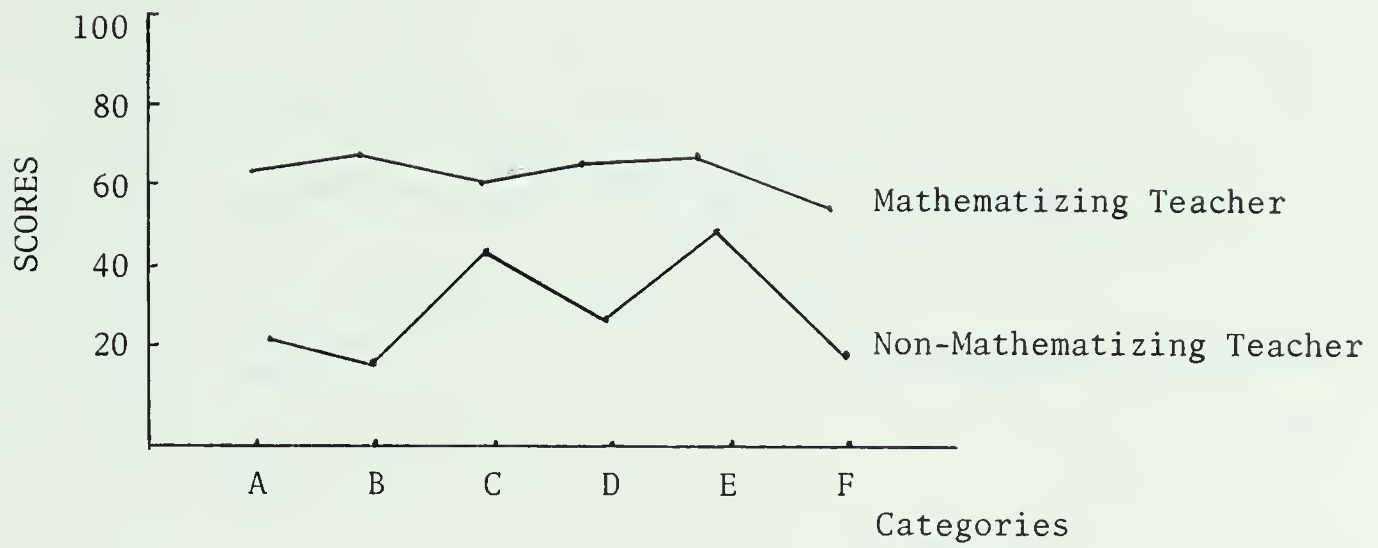

FIGURE 1

PROFILES OF THE TWO PILOT STUDY TEACHERS' CLASSROOM BEHAVIORS (BASED ON DATA FROM OBSERVER RATING SCALE ON THE FIRST OBSERVATION)

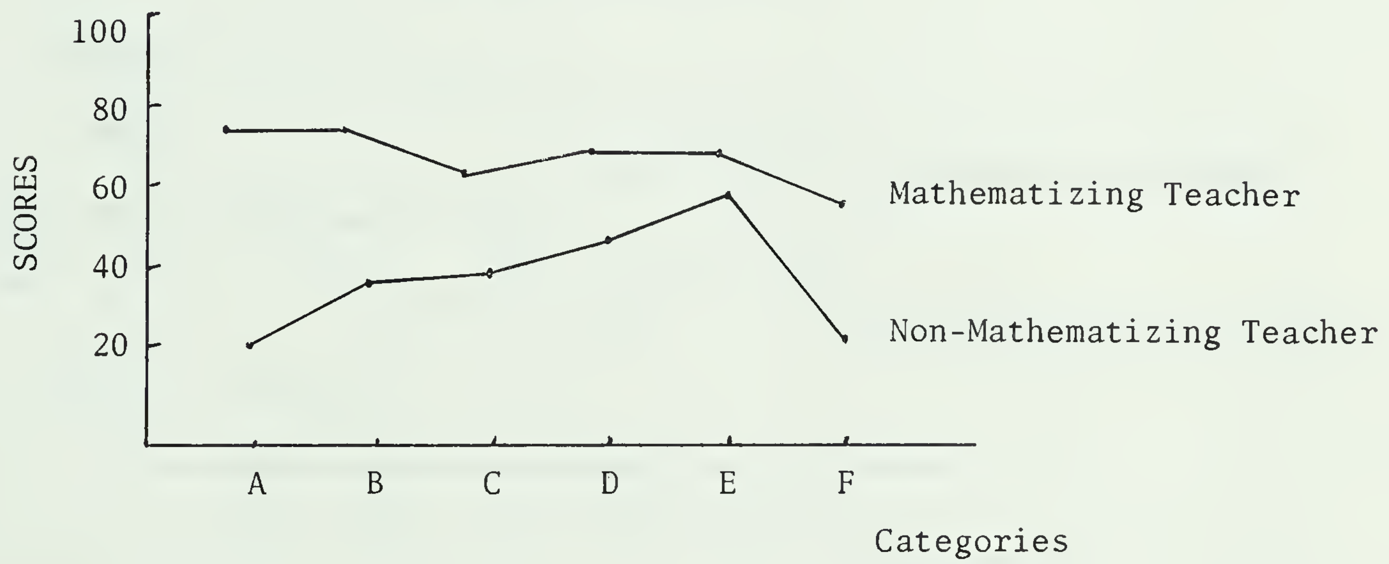

FIGURE 2

PROFILES OF THE TWO PILOT STUDY TEACHERS' CLASSROOM BEHAVIORS (BASED ON DATA FROM OBSERVER RATING SCALE ON THE SECOND OBSERVATION)

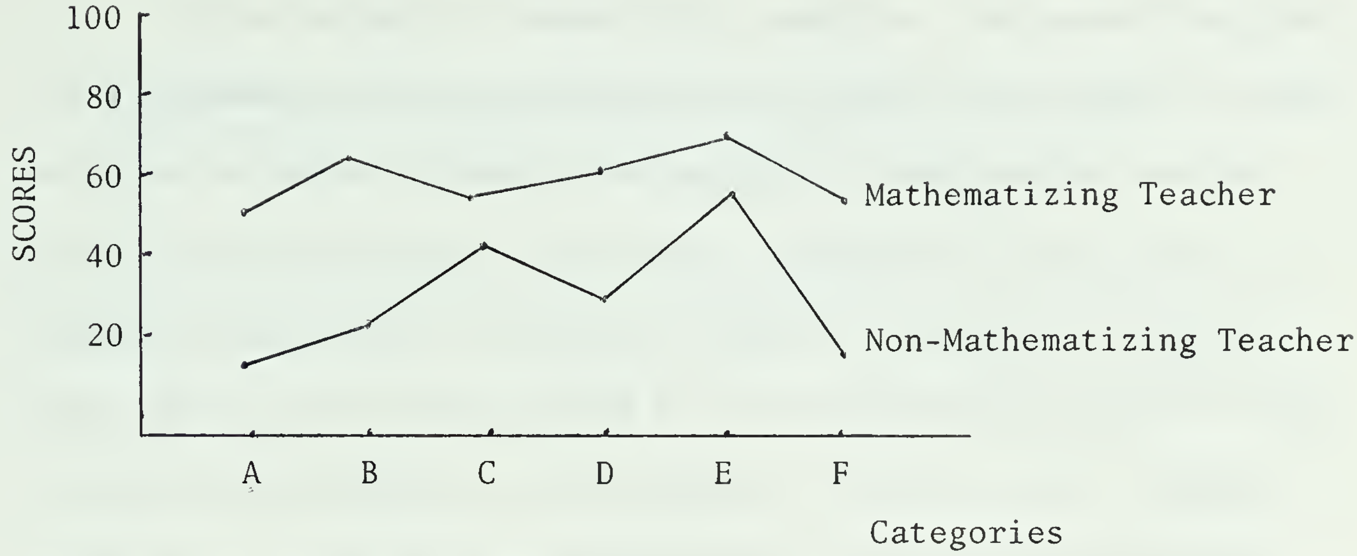

FIGURE 3

PROFILES OF THE TWO PILOT STUDY TEACHERS' CLASSROOM BEHAVIORS (BASED ON DATA FROM STUDENT INVENTORY ON FIRST COMPLETION)

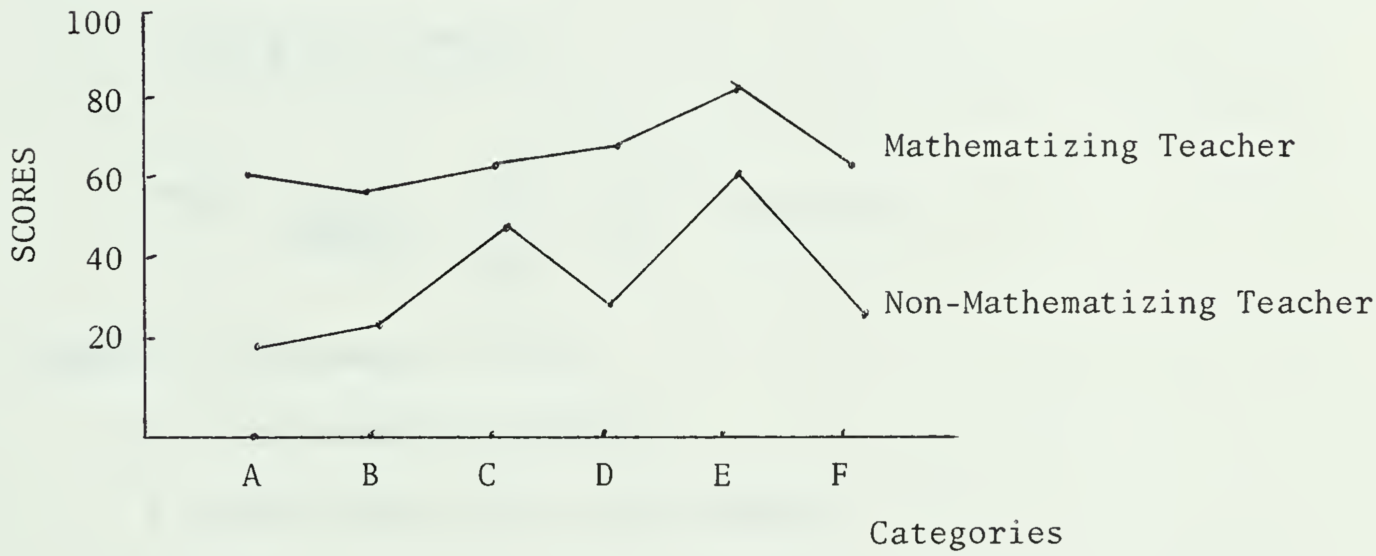

FIGURE 4

PROFILES OF THE TWO PILOT STUDY TEACHERS' CLASSROOM BEHAVIORS (BASED ON DATA FROM STUDENT INVENTORY ON SECOND COMPLETION)

In an attempt to answer this question, the scores from the two instruments for the mathematizing teachers were ranked for each category as shown in Table IV on page 69. Those scores obtained from the two instruments were ranked similarily for the non-mathematizing teacher as shown in Table V on page 70. Each of these sets of ranks were tested by the Friedman $\chi^2 r$ test[11] for significance of difference in rank orders. This test for k related samples was considered appropriate here since it is an over-all test of whether the size of the scores depends upon the conditions under which they were obtained.[12] Its power efficiency compares with the parametric F test.

Using the formula,

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (Rj)^2 - 3N(k+1)$$ [13]

where N = the number of rows

k = the number of columns

R_j = Sum of the ranks in jth column

$\sum_{j=1}^{k}$ directs one to sum the squares of the sums of ranks over all k conditions

The rank orders of the mathematizing teacher yielded a $\chi^2_r = 9.66$.

---

[11]Siegel, op. cit., p. 173.

[12]Siegel, op. cit., p. 166.

[13]Siegel, op. cit., p. 168.

TABLE IV

SCORES OBTAINED FROM THE OBSERVER RATING SCALE AND THE STUDENT INVENTORY FOR THE MATHEMATIZING TEACHER IN THE PILOT STUDY RANKED ACROSS CATEGORIES

Category	1st Obs. Rating	2nd Obs. Rating	1st St. Inv.	2nd St. Inv.
A	2	4	1	3
B	3	4	1	2
C	1.5	1.5	3	4
D	2	4	1	3
E	1.5	1.5	3	4
F	2	4	1	3

$\chi^2_r = 9.66$ $\quad .05 > p > .02$

TABLE V

SCORES OBTAINED FROM THE OBSERVER RATING SCALE AND THE STUDENT INVENTORY FOR THE NON-MATHEMATIZING TEACHER IN THE PILOT STUDY RANKED ACROSS CATEGORIES

Category	First Obs. Rating	Second Obs. Rating	First St. Inv.	Second St. Inv.
A	3	4	1	2
B	1	4	2	3
C	2.5	1	2.5	4
D	1	4	2	3
E	1	4	3	2
F	1	4	3	2

$\chi^2_r = 9.78$ $\quad .05 > p > .02$

The probability of occurence of $\chi^2_r = 9.66$ under the null hypothesis of no difference of rank orders of the scores is $.05 > p > .02$. The rank orders of the scores for the non-mathematizing teacher yielded a $\chi^2_r = 9.78$ with the same probability of occurence under the null hypothesis of no difference in the rank orders of the scores.

Although the data from the Observer Rating Scale and the Student Inventory are not similar enough to be accepted at the .05 level of significance, they are similar enough to be accepted at the .02 level. This evidence indicates some promise for the assumption that the scores obtained from the two instruments could be combined into one composite rating.

Question 3. To what extent do the ranked scores from the Observer Rating Scale agree with the ranked scores from the Student Inventroy as to Within Teacher methods rated?

The Within Teacher method was taken to be represented by the rank orders of the six Teacher Categories obtained on a single observation. Tables VI and VII of the next page show the intercorrelation matrices of the ranked Teacher Categories for the Non-mathematizing and Mathematizing teacher.

The Spearman rank correlation coefficient r_S was calculated for each pair of scores for each teacher. This nonparametric correlation coefficient was chosen because it uses as much information of

TABLE VI

CORRELATION MATRIX (SPEARMAN r_S) OF RANKED SCORES FOR THE NON-MATHEMATIZING TEACHER IN THE PILOT STUDY

	Obs. Rating_1	Obs. Rating_2	St. Inv._1	St. Inv._2
Obs._1		.72	.81	.90#
Obs._2			.88#	.88#
St. Inv._1				.94#

#Indicates significance at the .05 level.

TABLE VII

CORRELATION MATRIX (SPEARMAN rS) OF RANKED SCORES FOR THE MATHEMATIZING TEACHER IN THE PILOT STUDY

	Obs. Rating_1	Obs. Rating_2	St. Inv._1	St. Inv._2
Obs._1		.81	.46	.06
Obs._2			.20	-.35
St. Inv._1				.75

the data as Kendall τ but is easier to calculate and its power efficiency is the same, 91 percent of that of the parametric coefficient Pearson r.[14]

An inspection of Table VI shows that the Spearman r_S correlations with only the first Observer Rating scores do not reach the .05 level of significance for the Non-mathematizing teacher. This could be due to the investigator's inexperience with the Observer Rating Scale. The relatively high level of agreement among the ranked scores for the Non-mathematizing teacher as indicated by the Spearman r_S coefficients seems to indicate that this teacher was internally consistent within the method he used. Furthermore, the students appeared to be highly agreed on what went on in the classroom.

An inspection of Table VII reveals that no Spearman r_S correlations reached the .05 level of significance although the ranks of the two Observer Rating Scale scores which correlated .81 and ranks of the Student Inventory scores correlated .75 are somewhat higher than the intercorrelations of ranks from the two instruments. This seems to indicate:

1. That the mathematizing teacher was not internally consistent with his methods. That is, there was a Within Methods variation in the lessons.
2. The students had not been exposed to the developing method sufficiently to be in agreement on what was happening in the classroom.

---

[14]Siegel, op. cit., p. 239.

It appears from the intercorrelations that if the Within Teacher's method is stable for some time, the scores obtained on either instrument are highly related. However, if the Within Teacher's method varies abruptly, the scores from the Student Inventory do not reflect this variation as precisely as those from the Observer Rating Scale.

Conclusions

The pilot study was conducted to test the two instruments in two different teaching situations. An attempt was made to find out whether the scores from the instruments discriminated between the methods observed, whether the scores for each teacher were sufficiently similar to warrant combination, and whether the instruments required further revision. The results of the analysis of the data led this investigator to draw the following conclusions:

1. Scores, when ranked for each category, indicated that the two teachers used different methods.
2. Scores obtained for each teacher from the two instruments indicated some promise of agreement for the purpose of combination.
3. The greater variation in the rank orders of scores for the mathematizing teacher suggests that ratings of teacher methods should be carried out as near to the end of the experimental period as possible for greatest level of agreement.
4. Several items in each instrument required further rephrasing for clarification of intent.

III. VALIDITY AND RELIABILITY OF THE TWO INSTRUMENTS

Validity

The validity of the two instruments was established in the following ways. First, since the definitions of each Teacher Category describe the kinds of things teachers do in the mathematics classroom, the instruments have face validity. Second, since the Teacher Categories used attempt to cover the total activities of the lesson, the criterion of inclusiveness of categories is met and provides further evidence of the validity of the two instruments. Third, a jury of two university professors and five graduate students, very familiar with the two methods and experienced in teaching, judged each of the items as highly expository or highly mathematizing to establish the direction of the weighting scale and provide evidence of content validity. Fourth, Tables II and III from the pilot study show that none of the Teacher Category scores of the mathematizing teacher overlap with scores for the non-mathematizing teacher which indicates that the instruments developed in this study do discriminate between two different kinds of teachers in terms of teaching behaviors and thus provide further evidence for the validity of the instruments.

Reliability

Evidence of the reliability of the two instruments was obtained through the test retest method in the pilot study. Since each

of the six Teacher Categories of the instruments attempt to identify different things, the instruments satisfy Guilford's definition of a heterogeneous test. Hence according to Guilford, the only meaningful estimate of reliability is of the retest variety.[15] The results of the test retest gave a Spearman r_S of .81 for the Observer Rating Scale in the mathematizing class and an r_S of .72 in the non-mathematizing class. The Student Inventory yielded an r_S of .75 for the mathematizing class and an r_S of .94 for the non-mathematizing class. Although only an r_S of .94 is significant at the .05 level, tests of significance are not necessarily appropriate for one expects some variation in teaching methods as well as variations in the ratings between the test and retest,[16] which in this case was a period of three weeks. The Spearman correlation coefficients are deemed sufficiently high to indicate a degree of reliability of the instruments.

IV. SUMMARY

The present chapter was devoted to a detailed description of the rationale underlying the construction of the Observer Rating Scale of Teacher Behaviors and the Student Inventory of Teacher

---

[15]Guilford, op. cit., p. 450.

[16]Guilford, op. cit., p. 299.

Behavior. The six Teacher Category Indexes were described operationally and the scoring procedure used for each instrument was explained in detail.

Results from the pilot study were analyzed to lend support for the main assumptions under which the instruments were developed. The results of the pilot study presented some evidence for the validity and reliability of the two instruments developed to obtain measures of teacher classroom behaviors in terms of the methods used.

CHAPTER V

THE RESULTS AND INTERPRETATIONS OF THE INVESTIGATION

This chapter contains the results and interpretations of the statistical tests used in testing the null hypotheses of this study. Variations in Teacher Category Index scores were taken to reflect variations within teachers' methods. Variations between scores for each Teacher Category Index were taken to reflect variations between teachers' methods.

Analysis of the data was carried out utilizing nonparametric tests since the scores obtained were considered to have relative meaning only. To use as much information from the obtained scores as possible, the scores were translated into ranks. Statistical tests were carried out on differences in teachers' ranks so obtained. Data was cast in row by column tables so that tests carried out on the differences in Teacher Category Index ranks across columns were interpreted as tests on _Between Teachers' Methods_. Tests carried out on the differences in Teacher Category Index ranks down columns were interpreted as tests on _Within Teacher's Methods_.

Part I of this chapter contains the results and interpretations of the statistical tests on the ranks obtained before the in-service treatment which are used in testing null hypotheses 1 and 2. The results and interpretations of the tests on the during-treatment ranks which are used in testing null hypotheses 3 through 6 are

presented in Part II. Part III contains the results and interpretations of tests on the ranks obtained after the experimental period which are used in testing null hypotheses 7 and 8.

PART I: PRE-TREATMENT RESULTS AND INTERPRETATIONS

The before in-service treatment raw scores from the Observer Rating Scale and the Student Inventory are tabulated in Appendixes F.1 and F.2, respectively. For the purpose of testing null hypotheses 1 and 2, these scores were translated into ranks as described following the statement of each of the hypotheses.

Null Hypothesis 1

There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes before the in-service treatment.

Two tests were carried out on each teacher's methods ranks as indicated by scores from the Observer Rating Scale, and two were carried out on each teacher's methods ranks as indicated by scores from the Student Inventory. These tests together with the order in which they were carried out are indicated in the following table:

TABLE VIII

SCHEDULE OF STATISTICAL TESTS CARRIED OUT TO TEST HYPOTHESIS 1

Source	Test On	Statistical Test	Ranks
Obs. Rating	Between Methods	Friedman $\chi^2 r$	Between lessons
Scale	Within Methods	Kendall's W	Within lessons
St.	Between Methods	Wilcoxon T	Between lessons
Inventory	Within Methods	Spearman r_s	Within lessons

1. The Friedman $\chi^2 r$[1] test of significance of differences between rank orders of related samples was carried out on the ranks for each teacher as shown in Table IX, on the following page.

Using the formula,

$$\chi^2 r = \frac{12}{Nk\ (k + 1)} \sum (R_j)^2 - 3N(k + 1)$$

N = number of rows (in this case 6)

k = number of columns (in this case 4),

a $\chi^2 r \geq 7.82$ is needed to reject the hypothesis of no difference between ranks at the .05 level of significance since $\chi^2 r$ is distributed as chi-square with k - 1 degrees of freedom.

An inspection of Table IX shows that in each case, $\chi^2 r \geq 7.82$. This may be interpreted as indicating that the variations in rank

---

[1]Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, (Toronto: McGraw-Hill Book Company, 1956), pp. 166-173.

TABLE IX

PRE-TREATMENT CATEGORY RANKS FOR TEACHERS FROM OBSERVER RATING SCALE

	Teacher																											
	1				2				3				4				5				6				7			
	Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp	
Category	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	3	1	3	3	3	1	3	3	2	2	4	2	3	1	3	3	1.5	1.5	4	3	3.5	3.5	1.5	1.5	2	4	2	2
B	2	3	4	1	3	1	2	4	2	2	2	4	3.5	1	3.5	2	3.5	1.5	1.5	3.5	2.5	2.5	1	4	4	1.5	1.5	3
C	1.5	3.5	3.5	1.5	4	2	1	3	1.5	1.5	3	4	4	3	1	2	2.5	4	1	2.5	4	1	2.5	2.5	1	2.5	2.5	4
D	1.5	1.5	3	4	1.5	1.5	3.5	3.5	3	2	4	1	2	1	3.5	3.5	1.5	3	4	1.5	2.5	2.5	4	1	4	1	2.5	2.5
E	1.5	1.5	4	3	3.5	2	3.5	1	4	2.5	2.5	1	3.5	2	1	3.5	3	4	1.5	1.5	3	2	1	4	3.5	2	3.5	1
F	1	2	3.5	3.5	1	2	3	4	1	2.5	2.5	4	2.5	1	4	2.5	3	3	3	1	3	1	3	3	3	4	1.5	1.5
χ^2_r	6.4				4.4				1.8				5.1				0.8				2.3				0.9			
Prob. under H_o	$.10 > p > .05$				$.30 > p > .20$				$.70 > p > .50$				$.20 > p > .10$				$.90 > p > .80$				$.70 > p > .50$				$.90 > p > .80$			

orders of each Category for each of the teachers could have occurred by chance at that level of significance.

Hypothesis 1 is accepted for each of teacher's between methods ranks as indicated by Teacher Category Index scores from the Observer Rating Scale.

2. Kendall's Coefficient of Concordance W[2] was used to test the significance of the differences of rank orders when the Teacher Category Indexes were ranked for each lesson. Table X on the next page shows the rank orders on which Kendall's W was computed and the obtained Kendall's W.

Using the formula, $W = \dfrac{12\,S}{k^2(N^3 - N) - 4\sum T}$

where S = sum of the squared deviations from the mean of the rank sums.

N = number of rows (in this case 6)

k = number of columns (in this case 4)

T = the correction factor for tied ranks, a $W < .51$ is required to reject the hypothesis of no difference in rank orders at the .05 level of significance.

---

[2] Siegel, *op. cit.*, pp. 229-239.

TABLE X

PRE-TREATMENT TEACHER METHODS RANKS FROM OBSERVER RATING SCALE

	Teacher																											
	1				2				3				4				5				6				7			
	Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp	
Category	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	5.5	1.5	3	4	2	1	1.5	1.5	2	3.5	1.5	1	1.5	3	1.5	2	1	1	5	2	3	4	2	1	1	2.5	2.5	2
B	3.5	4	5	1	4.5	2.5	4.5	6	3	2	3	3.5	3	1.5	4	2	3.5	2.5	1.5	4	1	2	1	2.5	3.5	1	2.5	3.5
C	3.5	4	3	2	6	4.5	1.5	4.5	4	3.5	4.5	6	6	6	4	4.5	5	5	3.5	6	4.5	3	4	4.5	3.5	6	5	6
D	2	1.5	1	4	3	4.5	4.5	4.5	5	5	6	2	4.5	4	6	6	3.5	4	6	4	4.5	5	6	4.5	5	2.5	4	5
E	5.5	4	6	6	4.5	6	6	3	6	6	4.5	5	4.5	5	1.5	4.5	6	6	3.5	4	6	6	5	6	6	4.5	6	3.5
F	1	3	3	4	1	2.5	3	1.5	1	1	1.5	3.5	1.5	1.5	4	2	2	2.5	1.5	1	2	1	3	2.5	2	4.5	1	1
Kendall's W	.31				.66				.65				.63				.58				.82				.60			
Prob. under H_o	$p > .05$				$p < .05$				$p < .01$				$p < .05$				$p < .05$				$p < .01$				$p < .05$			

An inspection of Table X reveals that all teachers but teacher 1 have obtained $W > .51$. This may be interpreted as all teachers, with one exception, exhibited stability within methods of the lessons observed before the in-service treatment. Teacher 1 exhibited a degree of variation within methods of the lessons observed before the in-service treatment.

In summary, the results of tests on the before in-service treatment ranks obtained from the Observer Rating Scale indicate that all teachers exhibited stability between methods of lessons observed and that all but one teacher exhibited stability within methods of lessons observed.

3. The Wilcoxon Matched-Pairs Signed-Ranks[3] test was used to test the two sets of ranks obtained for each teacher for stability between methods for each teacher. This test determines whether a set of matched-pairs of ranks could have resulted from two different conditions by giving more weight to a pair which shows a large difference than to a pair which shows a small difference.[4] Table XI on the next page shows the matched pairs on which

---

[3]Siegel, op. cit., pp. 75-83.

[4]Ibid.

TABLE XI

WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST ON
PRE-TREATMENT SCORES OBTAINED BY TEACHERS ON STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Cat.	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	24	29	19	14	20	21	18	23	22	26	25	20	21	20
B	34	32	25	24	24	30	28	24	28	36	26	21	29	29
C	53	48	45	46	46	54	50	39	52	47	42	34	43	45
D	41	34	33	33	47	38	41	37	38	43	36	34	43	34
E	43	42	50	47	59	51	48	52	69	71	60	58	53	50
F	32	33	24	18	20	21	19	21	24	24	22	26	30	21
T	1.5		1.5		7.5		8		3.5		5		2	
N	6		5		6		6		5		6		5	

For N = 6, a Wilcoxon T = 0 is significant at $p < .05$.

the Wilcoxon T statistic was computed. An example of the computation of the Wilcoxon T statistic is given in Appendix D.1.

Exact probabilities associated with the obtained Wilcoxon T statistic for small samples were not available to this investigator, however, the Wilcoxon T approximates a normal distribution for samples greater than 25.

Table XI shows that the obtained T for each teacher is greater than 0. The hypothesis that there is no significant difference between the matched pairs of Category ranks for each teacher's classes is accepted. The result of the Wilcoxon test may be interpreted as indicating that the methods on which the two sets of Student Inventory scores were based were similar for each teacher at the .05 level of significance.

4. Spearman Rank Correlations[5] were calculated on Teacher Categories ranked for each teacher's class to test for significance of differences in rank orders for within each teacher's methods. The r_s is a measure of association between two sets of ranks. Table XII, on the following page, gives a summary of the obtained Spearman r_s calculated with the corresponding levels of probabilities of association.

---

[5]Siegel, *op. cit.*, pp. 202-213.

TABLE XII

SPEARMAN RANK CORRELATIONS CALCULATED ON PRE-TREATMENT TEACHER METHODS RANKS FROM STUDENT INVENTORY

Category	Teacher 1		2		3		4		5		6		7	
	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	1	1	1	1	1.5	1.5	1	2	1	2	2	1	1	1
B	3	2	3	3	3	3	3	3	3	3	3	3	2	3
C	6	6	5	5	4	6	6	5	5	5	5	4.5	4.5	5
D	4	4	4	4	5	4	4	4	4	4	4	4.5	4.5	4
E	5	5	6	6	6	5	5	6	6	6	6	6	6	6
F	2	3	2	2	1.5	1.5	2	1	2	1	1	2	3	2
Spearman r_s	.94		1.0		.83		.89		.94		.93		.93	
Probability Under H_o	$p < .05$		$p < .01$		$p < .05$		$p < .05$		$p < .05$		$p < .05$		$p < .05$	

With N = 6 an r_s = .829 is required for a one-tailed test of significant agreement of rank orders at the .05 level. An inspection of Table XII shows that in each case the obtained r_s exceeds that level. As a result, the hypothesis that there is no significant difference within a teacher's methods ranks as indicated by the Teacher Category Index scores obtained before the in-service treatment on the Student Inventory is accepted for each teacher.

In summary, tests on ranks obtained before the in-service treatment on the Student Inventory indicated no significant differences between a teacher's methods and no significant difference within a teacher's methods for each teacher.

Summary

Data from the two instruments for the teachers before the in-service treatment was converted to ranks in testing Hypothesis 1. Tests on ranks obtained from the Observer Rating Scale revealed that Teacher 1 exhibited significant variations within the methods used in the two classes. No significant variations were brought out from the tests between methods for each of the teachers. Tests on ranks obtained from the Student Inventory revealed no significant variations within methods or between methods for each of the teachers. Hypothesis 1 is accepted on this basis for all teachers but Teacher 1.

Null Hypothesis 2

There is no significant difference between teachers' methods ranks as indicated by the Teacher Category Index scores before the in-service treatment.

To test this hypothesis, means of the Teacher Category Indexes were calculated for the scores of the Observer Rating Scale and for those of the Student Inventory. Appendix D.2 contains the means of the Category Index scores so obtained.

The Friedman $\chi^2 r$ statistic was calculated on the ranks to tests for significance of differences between teachers' methods ranks as indicated by the Teacher Category mean scores. For this test, mean scores for each Category Index were ranked across teachers as shown in Tables XIII (a) and (b).

With N = 6, the number of categories, and k = 7, the number of teachers, a $\chi^2 r$ = 12.59 is required to reject the hypothesis of no significant difference among the rank orders at the .05 level.

Table XIII (a) shows that Category Index ranks obtained from the Observer Rating scores yielded a $\chi^2 r$ = 12.1, while Table XIII (b) shows that those from the Student Inventory scores yielded a $\chi^2 r$ = 14.6. At the .05 level of significance, hypothesis 2 is accepted on the basis of the rank orders indicated by scores from the Observer Rating Scale. Hypothesis 2 is rejected on this basis by scores from the Student Inventory.

TABLE XIII (a)

TEACHERS' RANKS FOR CATEGORIES BASED ON PRE-TREATMENT CATEGORY MEANS FROM THE OBSERVER RATING SCALE

	Teacher						
Category	1	2	3	4	5	6	7
A	7	1	2	3	4.5	6	4.5
B	6	3.5	3.5	1.5	7	1.5	5
C	2	1	5	6	3.5	3.5	7
D	2	1	5	6	3.5	7	3.5
E	4	1	6	2	4	7	4
F	7	1	2	3.5	3.5	6	5

$N = 6,\ k = 7 \quad \chi^2 r = 12.1 \quad .05 < p < .07$

TABLE XIII (b)

TEACHERS' RANKS FOR CATEGORIES BASED ON PRE-TREATMENT CATEGORY MEANS FROM THE STUDENT INVENTORY

	Teacher						
Category	1	2	3	4	5	6	7
A	7	1	3	3	6	5	3
B	7	1	3.5	2	6	3.5	5
C	7	4	5.5	3	5.5	1	2
D	3	1	7	4.5	6	2	4.5
E	1	2	5	3	7	6	4
F	7	2.5	2.5	1	4.5	4.5	6

$N = 6,\ k = 7 \quad \chi^2 r = 14.6 \quad p < .05$

Teachers were ranked for each Category Index as indicated by mean scores from each instrument. The Friedman $\chi^2 r$ statistic was calculated on each set of rank orders so obtained to test for significance of differences. The results of the Friedman test on the Observer Rating Scale scores yield a significant difference at the .07 level between the teachers' methods. The Student Inventory scores indicated, with the Friedman test, that there was a significant difference between the teachers' methods at the .05 level.

PART II: DURING-TREATMENT RESULTS AND INTERPRETATIONS

Null Hypothesis 3

There is no significant difference in a teacher's methods ranks as indicated by the Teacher Category Index scores for the two classes during the experiment.

Observer Rating Scale scores obtained on the two treatment methods for each teacher during the experimental period are presented in Appendix F.3. Those scores obtained from the Student Inventory during the same period are given in Appendix F.4. The averaged Observer Rating scores for each treatment Teacher Category Index are given in Table XIV, page 92.

The Wilcoxon T statistic was calculated on the ranks of difference scores for each teacher to test for significance of differences between pairs of scores. The scores from the two treatment methods were significantly different if the obtained Wilcoxon

TABLE XIV

WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST ON DURING EXPERIMENT CATEGORY MEANS FOR TREATMENT CLASSES BY TEACHERS FROM OBSERVER RATING SCALE

Category	Teacher													
	1		2		3		4		5		6		7	
	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	55	25	57.5	27.5	62.5	25	40	35	60	30	57.5	30	57.5	27.5
B	67.5	32.5	65	35	62.5	35	47.5	35	72.5	35	62.5	35	62.5	37.5
C	55	40	62.5	42.5	57.5	42.5	50	52.5	62.5	42.5	55	42.5	62.5	50
D	67.5	37.5	75	37.5	80	40	52.5	45	72.5	37.5	67.5	42.5	72.5	50
E	70	42.5	62.5	42.5	80	50	55	55	75	50	67.5	52.5	72.5	55
F	57.5	30	50	27.5	52.5	27.5	40	27.5	60	30	45	35	47.5	27.5
Wilcoxon T	0		0		0		1		0		0		0	
Prob.	$p < .01$		$p < .01$		$p < .01$		$.05 < p < .10$		$p < .01$		$p < .01$		$p < .01$	

TABLE XV

WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST ON TEACHERS' DURING-EXPERIMENT SCORES FROM STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Category	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	54	28	68	20	62	24	30	37	56	31	64	23	57	18
B	63	27	67	26	60	31	39	34	66	38	66	35	62	28
C	60	49	64	44	59	53	54	54	63	48	57	39	62	44
D	62	39	80	35	81	42	54	50	68	44	74	42	70	39
E	71	45	79	55	82	63	59	59	85	75	77	62	74	57
F	52	31	59	20	49	26	31	28	55	34	54	30	49	23
Wilcoxon T	0		0		0		4		0		0		0	
Prob.	$p < .05$		$p < .05$		$p < .05$		$p > .20$		$p < .05$		$p < .05$		$p < .05$	

T = 0. Tables XIV and XV both show that the Category Index scores for all teachers but Teacher 4 yielded a Wilcoxon T = 0, as obtained from the Observer Rating Scale and the Student Inventory. The results of the Wilcoxon T test indicate that scores from the two instruments reject Null hypothesis 3 for all teachers but Teacher 4.

A further inspection of the two tables reveals that all scores from the two instruments for the teachers' Mathematizing classes were higher than those for their Expository classes with the exception of Teacher 4. Teacher 4's scores are the same in Category E for both treatment classes. Table XIV shows that in Category C, scores obtained from the Observer Rating Scale are higher for the Expository class than for the Mathematizing class. In Table XV, scores obtained for that category from the Student Inventory are the same for each treatment class for that teacher.

Null Hypothesis 4

There is no significant difference between teachers' methods ranks as indicated by the Teacher Category Index scores during the experiment.

This hypothesis was tested on teachers ranked on Category Indexes for the Mathematizing classes as shown in Tables XVI (a) and (b), and on teachers ranked on Category Indexes for the

TABLE XVI (a)

FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE ON TEACHERS' RANKS BASED ON DURING MATHEMATIZING TREATMENT CATEGORY MEANS FROM THE OBSERVER RATING SCALE

Category	Teacher 1	2	3	4	5	6	7
A	2	4	7	1	6	4	4
B	6	5	3	1	7	3	3
C	2.5	6	4	1	6	2.5	6
D	2.5	6	7	1	4.5	2.5	4.5
E	4	2	7	1	6	3	5
F	6	4	5	1	7	2	3

$\chi^2_r = 22.2$ $p < .01$

TABLE XVI (b)

FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE ON TEACHERS' RANKS BASED ON DURING MATHEMATIZING TREATMENT CATEGORY SCORES FROM THE STUDENT INVENTORY

Category	Teacher 1	2	3	4	5	6	7
A	2	7	5	1	3	6	4
B	4	7	2	1	5.5	5.5	3
C	4	7	3	1	6	2	5
D	2	6	7	1	3	5	4
E	2	5	6	1	7	4	3
F	4	7	2.5	1	6	5	2.5

$\chi^2_r = 23.1$ $p < .01$

Expository classes shown in Tables XVII (a) and (b).

The Friedman $\chi^2 r$ statistic was calculated on each set of ranks to test for significance of differences among rank orders. With N = 6, the number of categories, and k = 7, the number of teachers, a $x^2 r \geq 12.59$ is required to reject the hypothesis of no difference in the rank orders at the .05 level of significance.

The Friedman test on the teachers' ranks determined by during-treatment scores obtained from the Observer Rating Scale for the Expository classes indicates that the rank orders are not significantly different at the .05 level. The Friedman test on the teachers' ranks determined by the during-treatment scores obtained from the same source for the Mathematizing classes indicates that the rank orders are significantly different at that level. The Friedman tests on the teachers' ranks obtained from the Student Inventory indicate that the teachers' ranks are significantly different for both the Mathematizing and the Expository classes.

In summary, tests on teachers' ranks determined by during-treatment scores from the two instruments revealed that there was a significant difference between teachers' methods in their Mathematizing classes during the experimental period. Data from the two instruments does not appear to agree as to whether a significant difference existed between teachers in the methods which they used in their Expository classes during the experimental period. It appears from this that the extent to which the teachers

TABLE XVII (a)

FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE ON TEACHERS' RANKS BASED ON DURING EXPOSITORY TREATMENT CATEGORY MEANS FROM THE OBSERVER RATING SCALE

	Teacher						
Category	1	2	3	4	5	6	7
A	1.5	3.5	1.5	7	4.5	4.5	3.5
B	1	4	4	4	4	4	7
C	1	3.5	3.5	7	3.5	3.5	6
D	2	2	4	6	2	5	7
E	1.5	1.5	3.5	6.5	3.5	5	6.5
F	5.5	2.5	2.5	2.5	5.5	7	2.5

$\chi^2 r = 10.3 \quad p < .05$

TABLE XVII (b)

FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE ON TEACHERS' RANKS BASED ON DURING EXPOSITORY TREATMENT CATEGORY SCORES FROM THE STUDENT INVENTORY

	Teacher						
Category	1	2	3	4	5	6	7
A	5	2	4	7	6	3	1
B	2	1	4	5	7	6	3
C	5	2.5	6	7	4	1	2.5
D	2.5	1	4.5	7	6	4.5	2.5
E	1	2	6	4	7	5	3
F	6	1	3	4	7	5	2

$\chi^2 r = 20.5 \quad p < .01$

were able to translate the theoretical Mathematizing model into classroom instructional behaviors varied significantly for both the outside observer and the students to recognize. On the other hand, the variations in the Expository methods were not great enough for the observer to recognize.

Null Hypothesis 5

During the experiment, there is no significant difference between ranks of teachers when ranks are determined by:

a) age and flexibility of methods;
b) years of teaching experience and flexibility of methods;
c) number of times they had taught Mathematics 20 and flexibility of methods;
d) number of university mathematics courses taken and flexibility of methods;
e) number of university degrees held and flexibility of methods;
f) percentile scores obtained on the Minnesota Teacher Attitude Inventory and flexibility of methods.

Flexibility of methods was defined as the sum of the deviations of each Method Index score during the experiment from the Method Index score before the experiment. One flexibility of methods score was calculated for each teacher on the basis of the data obrained from the Observer Rating Scale and another on the basis of data obtained from the Student Inventory. The flexibility of methods for each teacher along with the procedure used are given in Appendix E.

To test Hypothesis 5, teachers were ranked on the selected characteristics, and on each set of flexibility of methods. Spearman r_s coefficients were calculated for set of ranks of characteristics paired with each set of ranks of flexibility of methods for the teachers as shown in Table XVIII. For N = 7, an $r_s = .714$ is required for significance at the .05 level.

A study of Table XVIII reveals that no coefficients reached the required level of significance. That is, the obtained r_s brings out no significant differences in the ranks between flexibility of methods and selected teacher characteristics. It is interesting to note that the flexibility of methods ranks from both instruments appear to correlate negatively with the ranks of teachers determined by the number of times they had taught Mathematics 20 and the number of University degrees held, and positively with the percentile ranks obtained on the Minnesota Teachers Attitude Inventory.

Null Hypothesis 6

There is no significant difference in the ranks of flexibility of methods between sexes during the experiment.

The Mann-Whitney U test[7] was used to test whether there was a significant difference in the ranks of flexibility of methods between the female and male teachers during the experiment.

---

[7]Siegel, *op. cit.*, pp. 116-126.

TABLE XVIII

TEACHERS' RANKS ON SELECTED CHARACTERISTICS
AND ON FLEXIBILITY OF METHODS

Characteristics	Teacher 1	2	3	4	5	6	7	r_s with Flex(a)	r_s with Flex.(b)*
Age	1	4.5	3	4.5	2	7	6	-.69	.39
Yrs. of Teaching Experience	1	5	3.5	3.5	2	7	6	-.52	.58
Yrs. of H.S. Math Experience	1	5.5	3.5	3.5	2	7	5.5	-.41	.64
No. of Times Math. 20 Taught	1	5	3	7	3	3	6	-.47	-.30
No. of Univ. Math. Courses	3	4	1	2	5	6	7	-.39	.05
No. of Degrees	1.5	1.5	4.5	4.5	4.5	4.5	7	-.44	-.18
M.T.A.I. Rank	5	6	4	1	2	3	7	.46	.55
Flexibility of Method from Obs. Scale (a)	6.5	5	6.5	1	4	2	3		.33
Flex. of Method from St. Inv.(b)	2.5	7	5.5	1	2.5	5.5	4		

*Spearman Correlation Coefficients by pairs of ranks

$r_s = .714$

$p = .05$

Tables XIX (a) and (b) on the following page show the ranks of the two independent groups.

The formula: $U = n_1 n_2 + n_1 \frac{(n_1 + 1)}{2} - R_1$

$n_1 = 3$

$n_2 = 4$

was used in calculating the Mann-Whitney U statistic.

The Mann-Whitney U statistic in each case shows no significant difference in the ranks of flexibility of methods between the males and the females during the experiment.

PART III: POST-TREATMENT RESULTS AND INTERPRETATIONS

Null Hypothesis 7

There is no significant difference in a teacher's methods for classes before the experiment and after the experiment.

Post-treatment raw scores are presented in Appendixes F.5 and F.6. Each teacher's pre- and post- Category Index scores from each instrument were paired in the following ways:

1. Pre- and Post- Category Index scores for the Mathematizing classes given in Tables XX and XXI.
2. Pre- and Post- Category Index scores for the Expository classes as shown in Tables XXII and XXIII.
3. Post- Category Index scores for the Mathematizing and Expository classes as given in Tables XXIV and XXV.

TABLE XIX (a)

CALCULATION OF THE MANN-WHITNEY U ON FLEXIBILITY OF METHODS RANKS FOR SEX FROM OBSERVER RATING SCALE

Females	6.5	5	2		$R_1 = 13.5$	$U = 4.5$, $p = .421$
Males	5.5	1	2.5	4	$R_2 = 13$	

TABLE XIX (b)

CALCULATION OF THE MANN-WHITNEY U ON FLEXIBILITY OF METHODS RANKS FOR SEX FROM THE STUDENT INVENTORY

Females	2.5	7	5.5		$R_1 = 15$	$U = 3$, $p = .200$
Males	5.5	1	2.5	4	$R_2 = 13$	

TABLE XX

WILCOXON TEST ON PRE- POST- MATCHED-PAIRS SCORES
IN MATHEMATIZING CLASSES FROM OBSERVER RATING SCALE

Category	Teacher 1		2		3		4		5		6		7	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	30	45	12.5	20	25	37.5	22.5	27.5	20	50	35	25	27.5	42.5
B	32.5	40	22.5	32.5	25	37.5	22.5	42.5	27.5	60	25	50	30	47.5
C	32.5	40	27.5	30	30	50	42.5	35	37.5	47.5	35	32.5	37.5	62.5
D	25	42.5	20	30	37.5	45	30	50	32.5	55	40	47.5	35	45
E	35	50	27.5	42.5	42.5	60	32.5	35	47.5	57.5	47.5	40	40	67.5
F	25	37.5	12.5	20	17.5	42.5	20	22.5	25	32.5	25	25	32.5	45
Wilcoxon T	0*		0*		0*		4		0*		7.5		0*	
Prob.							$p > .20$				$p > .20$			

*Significant difference between scores at $p < .01$.

TABLE XXI

WILCOXON TEST ON PRE- POST- MATCHED-PAIRS SCORES IN MATHEMATIZING CLASSES FROM STUDENT INVENTORY

Category	Teacher 1		2		3		4		5		6		7	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	24	29	19	22	20	43	18	26	22	38	25	33	21	33
B	34	33	25	23	24	48	28	29	28	44	26	34	29	41
C	53	53	45	42	46	51	50	51	52	56	42	46	43	51
D	41	48	33	36	47	63	41	46	38	50	36	40	43	52
E	43	58	50	47	59	69	48	61	59	71	60	70	53	65
F	32	32	24	27	20	32	19	24	24	40	22	28	30	35
Wilcoxon T	1		8		0*		0*		0*		0*		0*	
Prob.	$p = .05$		$p > .20$											

*Significant difference between scores at $p < .01$.

TABLE XXII

WILCOXON TEST ON PRE- POST- MATCHED PAIRS SCORES IN EXPOSITORY CLASSES FROM OBSERVER RATING SCALE

	Teacher													
	1		2		3		4		5		6		7	
Category	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	35	37.5	15	20	20	37.5	25	30	32.5	42.5	25	22.5	25	25
B	32.5	45	30	30	27.5	57.5	27.5	37.5	27.5	42.5	30	35	27.5	35
C	32.5	45	20	30	42.5	42.5	32.5	40	32.5	45	35	40	42.5	40
D	32.5	45	25	27.5	35	47.5	45	40	37.5	47.5	40	40	35	45
E	42.5	57.5	25	32.5	37.5	52.5	30	57.5	30	60	47.5	40	37.5	52.5
F	35	25	17.5	15	25	42.5	27.5	25	27.5	35	30	20	20	30
Wilcoxon T	1		1.5		0		3.5		0		5		1	
Prob.	$p = .05$		$.10 > p > .05$		$p < .01$		$p > .20$		$p < .01$		$p > .20$		$p = .05$	

TABLE XXIII

WILCOXON TEST ON PRE- POST- MATCHED-PAIRS SCORES
IN EXPOSITORY CLASSES FROM STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Category	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
A	29	27	14	18	21	24	23	36	26	31	20	28	20	20
B	32	28	24	24	30	37	24	35	36	39	27	32	29	31
C	48	45	46	42	54	46	39	55	47	42	34	43	45	42
D	34	42	33	35	38	40	37	53	43	41	34	37	34	34
E	42	49	47	50	51	59	52	60	71	68	58	67	50	53
F	33	28	18	17	21	24	21	33	24	33	26	32	21	26
Wilcoxon T	10		5.5		5.5		0*		8		0*		2.5	
Prob.	$p > .20$		$p > .20$		$p > .20$				$p > .20$				$.20 > p > .10$	

*Significant difference between scores at $p < .05$.

TABLE XXIV

POST-TREATMENT CATEGORY MEAN SCORES OBTAINED BY TEACHERS ON OBSERVER RATING SCALE

Category	Teacher 1		2		3		4		5		6		7	
	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	45	37.5	20	20	37.5	37.5	27.5	30	50	42.5	25	22.5	42.5	25
B	40	45	37.5	30	37.5	57.5	42.5	37.5	60	42.5	50	35	47.5	35
C	40	45	30	30	50	42.5	35	40	47.5	45	32.5	40	62.5	40
D	42.5	45	30	27.5	45	47.5	50	40	55	47.5	47.5	40	45	45
E	50	57.5	42.5	32.5	60	52.5	35	57.5	57.5	60	40	40	67.5	52.5
F	37.5	25	20	15	42.5	42.5	22.5	25	32.5	35	25	20	45	30
Wilcoxon T	10.5		0		5		8.5		4		2.5		0	
Prob.	$p > .20$		$p < .01$		$p > .20$		$p > .20$		$p > .20$		$.10 < p < .20$		$p < .01$	

TABLE XXV

POST-TREATMENT SCORES OBTAINED BY TEACHERS ON STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Category	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	29	27	22	18	43	24	26	36	38	31	33	28	33	20
B	33	28	23	24	48	37	29	35	44	39	34	32	41	31
C	53	45	42	42	51	46	51	55	56	42	46	43	51	42
D	48	42	36	35	63	40	46	53	50	41	40	37	52	34
E	58	49	47	50	69	59	61	60	71	68	70	67	65	53
F	32	28	27	17	32	24	24	33	40	33	28	32	35	26
Wilcoxon T	0		4.5		0		1		0		5		0	
Prob.	$p < .01$		$p > .20$		$p < .01$		$p = .05$		$p < .01$		$p > .20$		$p < 01$	

The Wilcoxon T statistic was computed for each matched-pair of signed-ranks so arranged, to test for the significance of the differences between the scores of each pair. For N = 6, a T = 0 is required for a significant difference between matched-pairs at the .05 level.

A study of the above tables reveals that the Wilcoxon T test based on data from each instrument brings out no significant difference in teacher 2's methods for his Expository classes before the experiment and after the experiment. The Wilcoxon T test based on data from each instrument shows a significant difference in teachers 3, 5, and 7 methods for the Mathematizing classes before and after the experiment. The Wilcoxon T test based on data from each instrument shows a significant difference in teacher 7's methods and no significant difference in teacher 6's methods when the Mathematizing class was paired with the Expository class after the experiment.

Hypothesis 7 was rejected for all teachers but 4 and 6 on the basis of data from the Observer Rating Scale. Hypothesis 7 was rejected for all teachers but Teacher 2 on the basis of data from the Student Inventory. The methods used by most teachers after the experiment were different from those they used before the experiment.

Null Hypothesis 8

There is no significant difference between teachers' methods for classes after the experiment.

Teachers ranks on Post-treatment Category Index scores in the Mathematizing classes are shown in Tables XXVI (a) and (b). Teachers ranks on Post-treatment Category Index scores in the Expository classes are given in Tables XXVII (a) and (b).

The Friedman $\chi^2 r$ statistic was computed to test for significance of differences between rank orders so obtained. With N = 6, the number of categories, and k = 7, the number of teachers, a $\chi^2 r \geq 12.59$ is required to reject the null hypothesis of no difference between rank orders at the .05 level of significance.

An inspection of the above tables shows that each of the obtained $\chi^2 r$ exceed 12.59. The $\chi^2 r$ test shows a significant difference exists between the ranks as indicated by the scores from each instrument.

Null Hypothesis 8 is rejected on this basis. This is, there is a significant difference between teachers' methods for classes after the experiment as indicated by the Teacher Category Index scores from both instruments.

SUMMARY

In testing each of the null hypotheses of this study, the scores were converted to ranks since the scores from each instrument have relative meaning only. Nonparametric statistical tests were carried out on teachers' ranks in testing for significant differences. Statistical decisions reached with respect to each of the null hypotheses are summarized in Table XXVIII.

TABLE XXVI (a)

TEACHERS' RANKS ON POST-TREATMENT CATEGORY MEAN SCORES FOR MATHEMATIZING CLASSES FROM OBSERVER RATING SCALE

	Teacher						
Category	1	2	3	4	5	6	7
A	6	1	4	3	7	2	5
B	3	1.5	1.5	4	7	6	5
C	4	1	6	3	5	2	7
D	2	1	3.5	6	7	5	3.5
E	4	3	6	1	5	2	7
F	5	1	6	2	4	3	7

$N = 6 \quad k = 7 \quad \chi^2 r = 18.6 \quad p < .01$

TABLE XXVI (b)

TEACHERS' RANKS ON POST-TREATMENT CATEGORY MEAN SCORES FOR MATHEMATIZING CLASSES FROM THE STUDENT INVENTORY

	Teacher						
Category	1	2	3	4	5	6	7
A	3	1	7	2	6	4.5	4.5
B	3	1	7	2	6	4	5
C	6	1	4	4	7	2	4
D	4	1	7	3	5	2	6
E	2	1	5	3	7	6	4
F	4.5	2	4.5	1	7	3	6

$N = 6 \quad k = 7 \quad \chi^2 r = 25.5 \quad p < .01$

TABLE XXVII (a)

TEACHERS' RANKS ON POST-TREATMENT CATEGORY MEAN SCORES FOR EXPOSITORY CLASSES FROM THE OBSERVER RATING SCALE

Category	Teacher 1	2	3	4	5	6	7
A	5.5	1	5.5	4	7	2	3
B	6	1	7	4	5	2.5	2.5
C	6.5	1	5	3	6.5	3	3
D	4.5	1	6.5	2.5	6.5	2.5	4.5
E	5.5	1	3.5	5.5	7	2	3.5
F	3.5	1	7	3.5	6	2	5

$N = 6 \quad k = 7 \quad \chi^2 r = 23.1 \quad p < .01$

TABLE XXVII (b)

TEACHERS' RANKS ON POST-TREATMENT CATEGORY MEAN SCORES FOR EXPOSITORY CLASSES FROM THE STUDENT INVENTORY

Category	Teacher 1	2	3	4	5	6	7
A	4	1	3	7	6	5	2
B	2	1	6	5	7	4	3
C	5	2	6	7	2	4	2
D	6	2	4	7	5	3	1
E	1	2	4	5	7	6	3
F	4	1	2	6.5	6.5	5	3

$N = 6 \quad k = 7 \quad \chi^2 r = 21.8 \quad p < .01$

TABLE XXVIII

SUMMARY OF STATISTICAL DECISIONS REACHED ON EACH HYPOTHESIS

Hypothesis	Decision based on the Observer Rating Scale	Decision based on the Student Inventory
1	accept for all teachers but Teacher 1	accept
2	reject[#]	reject
3	accept for Teacher 4 only	accept for teacher 4 only
4	reject for Math. class accept for Exp. class	reject
5	accept	accept
6	accept	accept
7	reject for all teachers but 4 and 6	reject for all teachers but 2
8	reject	reject

[#]Decision based on .07 level of significance, all others based on .05 level.

As is indicated in Table XXVIII, probabilities tested on data from the two instruments did not lead to the same decision in each case. Further evidence of this discrepancy is presented in profiles of Method Indexes for Teachers shown in Figures 5 through 8 of the following pages.

An inspection of the profiles reveals that in each case during the experiment, teachers' methods as indicated by Method Indexes moved in the direction of the Mathematizing treatment model from the position occupied before the in-service. After the experiment, each teacher situated toward the Expository treatment model in the Mathematizing class while there was a variety of movements in the Expository classes.

The Post-Experiment profiles in each case appear to have a greater variation than those of the Pre-Inservice treatment. A Sign test[8] on deviations from the means of the Method Indexes supports this observation at .02 level for the Mathematizing classes from each instrument and at the .10 level for the Expository classes. This is also reflected in the tests of null hypotheses 2 and 8. Null hypothesis 2 was rejected at the .05 and .07 levels of significance while null hypothesis 8 was rejected at the .01 level of significance.

---

[8]Siegel, *op. cit.*, pp. 68-75.

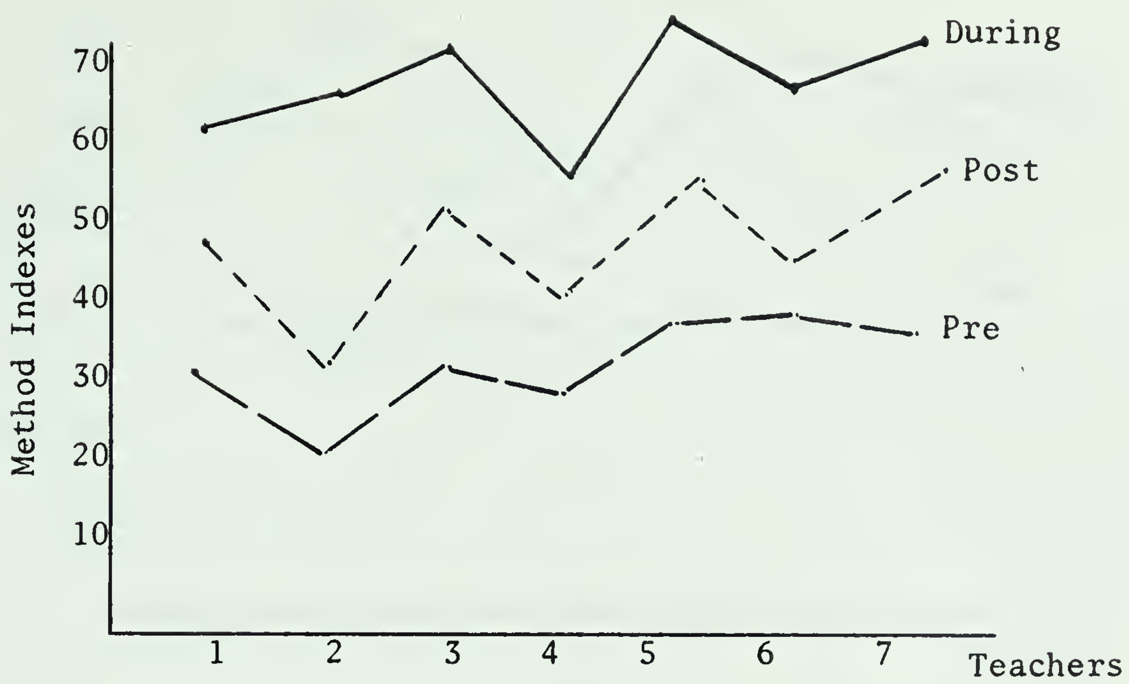

Figure 5

TEACHER METHOD INDEXES FOR MATHEMATIZING TREATMENT CLASSES
(BASED ON OBSERVER RATING PRE-POST-DURING- EXPERIMENT DATA)

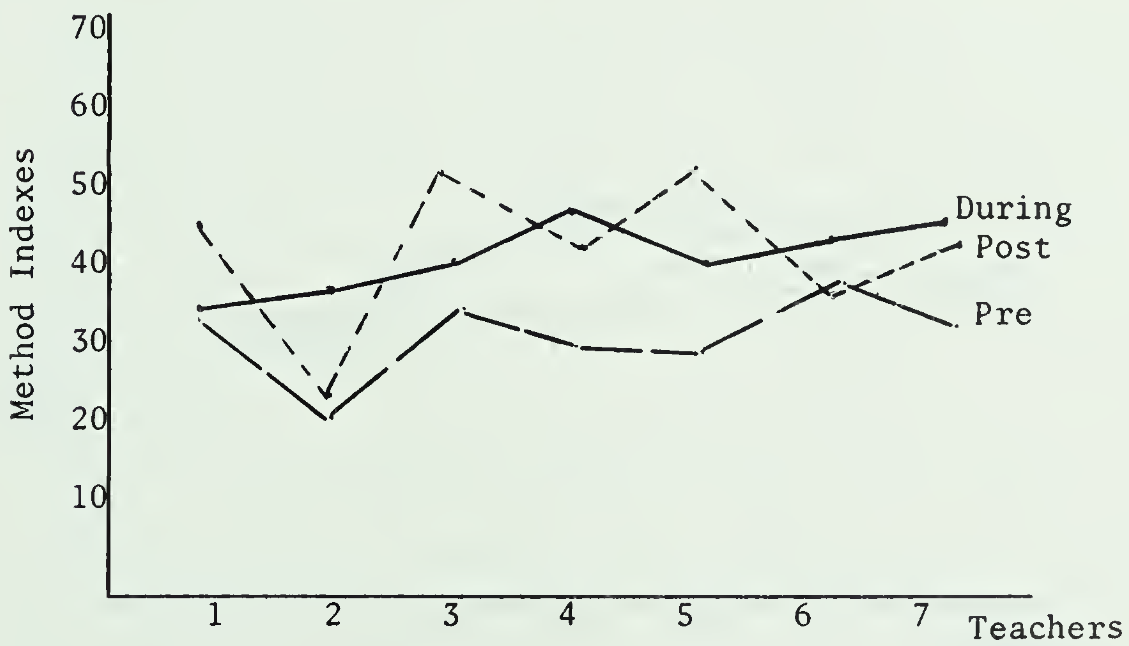

Figure 6

TEACHER METHOD INDEXES FOR EXPOSITORY TREATMENT CLASSES
(BASED ON OBSERVER RATING PRE-POST-DURING- EXPERIMENT DATA)

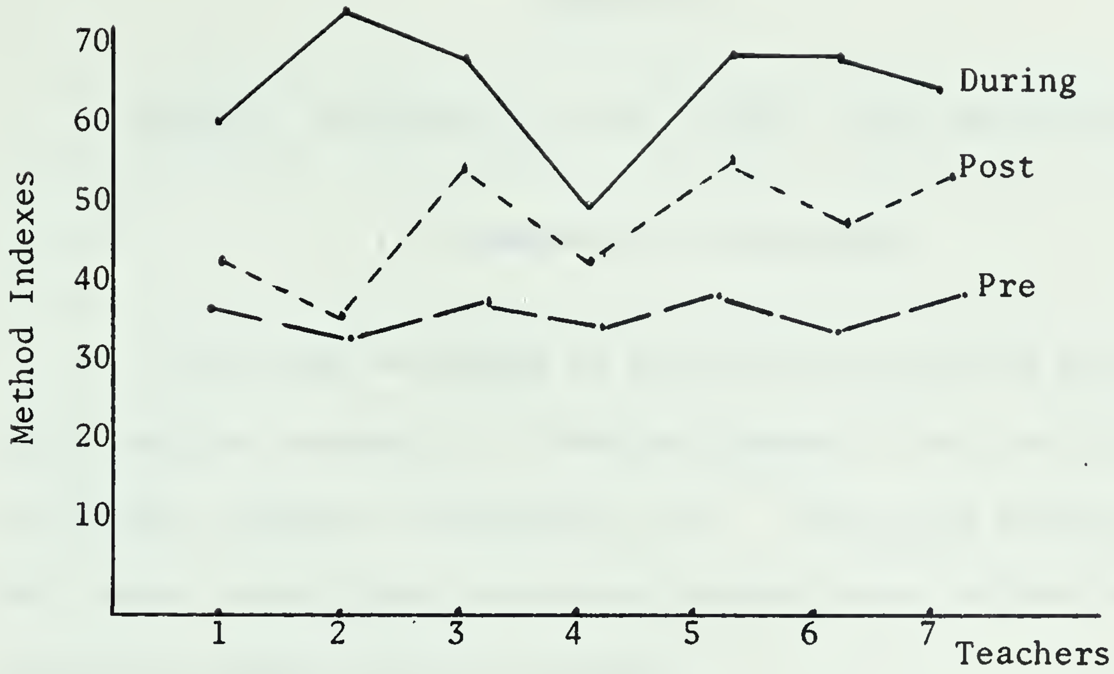

Figure 7

TEACHER METHOD INDEXES FOR MATHEMATIZING TREATMENT CLASSES
(BASED ON STUDENT INVENTORY PRE-POST-DURING EXPERIMENT DATA)

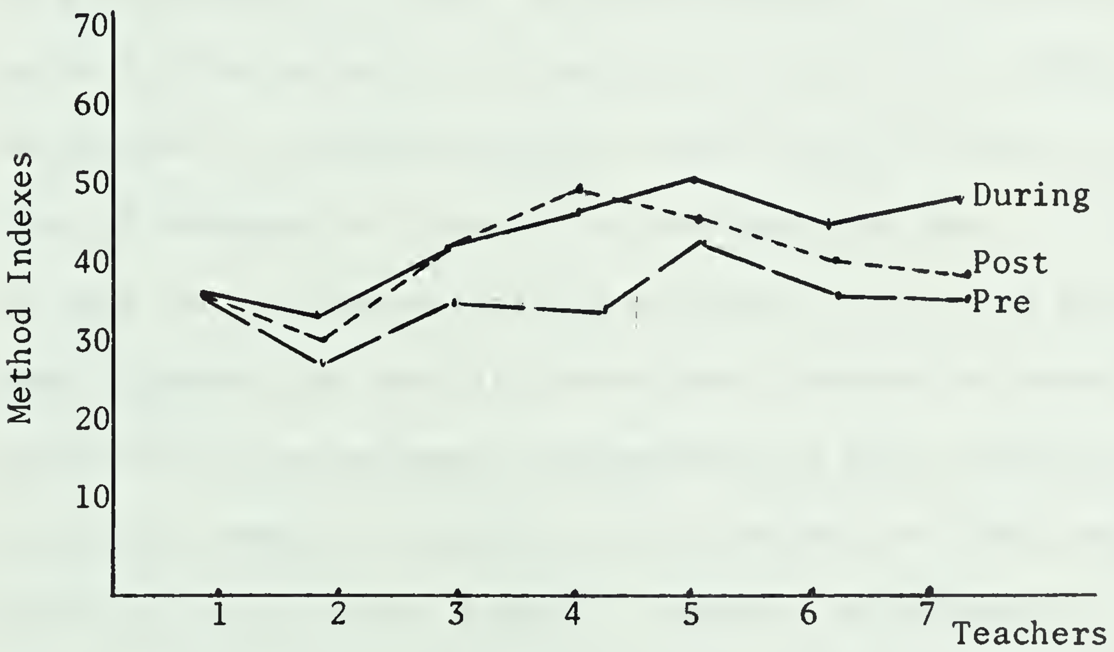

Figure 8

TEACHER METHOD INDEXES FOR EXPOSITORY TREATMENT CLASSES
(BASED ON STUDENT INVENTORY PRE-POST-DURING- EXPERIMENT DATA)

CHAPTER VI

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND LIMITATIONS

I. SUMMARY AND CONCLUSIONS

This study attempted to evaluate the relative potential of an in-service program for effecting change in the teaching methods of certain teachers of Mathematics 20. The second problem in this study investigated techniques through which objective data on teachers' methods could be obtained.

Each of the seven teachers involved in this study attended five two-hour in-service sessions held once every two weeks beginning in February of 1968. At these sessions, the teachers were given printed materials outlining the instructional behaviors to be followed in teaching specific topics from the units on quadratics of Mathematics 20 using the Mathematizing Mode. In the face-to-face lecture-discussion type sessions, the printed materials were discussed and parallels were drawn between the behaviors required for the Mathematizing method with those required for the Expository method of teaching. Each teacher was then required to teach two units on the quadratic function of Mathematics 20 to one class using the Mathematizing Mode and to the other class using the Expository Mode.

The review of the literature shows that a number of studies have investigated and analyzed classroom interaction patterns. Through the use of various interaction analyses, researchers have been able to measure teacher classroom verbal behaviors more scientifically than through the use of teacher or student opinionnaires. However, two problems seem to exist. First, the results of these interaction analyses do not seem to lend themselves to interpretations with respect to direct measures of methods of instruction the teacher is using. The methods observed must be inferred from the analyses. Second, although the instruments and techniques now available yield greater precision to the appraisal of teacher classroom behaviors empirically, the polarization of methods, however, still continues on a theoretical basis. Instruments develop to measure the extent to which teachers adhere to particular bi-polar methods have been based on the theoretical frameworks of the methods with which the individual researchers was concerned.

Although a number of researchers have investigated the relative potential of content in-service programs, a limited number of investigations have evaluated the relative potential of methods in-service programs for effecting change in the participants' teaching methods. Evaluations of the effects of in-service programs have been carried out more often in terms of student achievement or

teacher effectiveness than in terms of changes in teachers methods of instruction. Those studies that did attempt to evaluate the effects of in-service programs in terms of changes in teachers methods have found that generally teachers changed their methods somewhat as a result of participation in in-service education. However, most of the studies were based on results from teacher and student opinionnaires rather than on before and after the in-service ratings obtained by outside observers.

This study was designed to evaluate the relative potential of a particular methods in-service program. An outside observer as well as students rated teacher instructional behaviors before the in-service program, during the experimental treatment, and after the experimental period, in an attempt to identify significant changes.

Two instruments were developed in this study through which this investigator attempted to obtain measures of teaching methods. The instruments were based on six teacher behavioral dimensions called Teacher Category Indexes. Five related items were rated under each Teacher Category to yield a score for that category. The two treatment methods were considered bi-polar on the rating continuum of each Teacher Category. Variations in Teacher Category Index scores were hypothesized as indicating variations within the teaching methods. Variations between scores for each Teacher Category Index were hypothesized as indicating variations between teachers' methods.

The two instruments were tested in a pilot study before the main experiment. The results of the pilot study indicated that each instrument succeeded in discriminating difference within a teacher's methods and between the two teachers for methods.

Scores from the two instruments were translated into ranks for the purpose of testing for significant differences in methods. This was considered appropriate since statistical tests on rank orders appeared to be consistent with flexibility in methods as hypothesized. Nonparametric techniques were used in the analysis of the data so ranked. Probabilities were investigated for significance by tests considered appropriate.

The analysis of classroom interactions under the two treatment models, Mathematizing and Expository, revealed both stability within a teacher's methods and deviations between teachers' methods, before the in-service treatment. During the experimental period, all but one teacher, were found to be teaching the two classes using methods approximating the prescribed treatment models. One teacher was found to be teaching both classes with a method approximating the mean between the two treatment models. All teachers were found to have adjusted their teaching methods in the direction of the Mathematizing model after the experimental period. Four out of the seven teachers, however, were found to be sustaining significantly the methods they were using during the experimental period.

Specifically, the more important findings were:

1. There was no significant difference in the methods a teacher used in his classes prior to the in-service treatment.

2. There was a significant different between teachers' methods before the in-service treatment.

3. All teachers changed their teaching methods significantly during the experimental period.

4. All teachers, with one exception, taught each of the two treatment classes using two different methods during the experimental period. Teacher 4 used basically the same method in each treatment class.

5. The extent to which each teacher was able to vary his teaching methods was not significantly related to the selected teacher characteristics. A negative relationship was found between flexibility of methods and the number of times teachers had taught Mathematics 20, as well as flexibility of methods and the number of university degrees the teachers held. A positive relationship was found between flexibility of methods and permissiveness as indicated by the Minnesota Teacher Attitude Inventory.

6. There was no significant difference in the flexibility of teaching methods between the sexes.

7. Teachers 2, 4, and 6 used methods after the experiment similar to those they used before the in-service treatment. The other teachers used significantly different methods after the

experiment from those they used before the in-service treatment.

8. After the experimental period, the teachers exhibited a greater variation between the teaching methods they used than before the in-service treatment.

9. Although the two instruments used to gather data on teachers' methods were designed to parallel each other, the data obtained from the Observer Rating Scale was not always compatible with that obtained from the Student Inventory. Data from the two instruments agreed on those teacher behaviors which appeared to be constant over a period of time. They did not appear to agree on teacher behaviors which were in a state of flux.

Several conclusions are drawn from the findings of this study.

1. An in-service program designed to operationally define methods of instruction has potential for effecting change in the teaching methods of certain teachers.

2. The extent to which the teachers are able to translate theoretical models of instruction into classroom practices appears to be related positively to the teachers' permissiveness, and negatively to the number of degrees and the number of times the teacher had taught Mathematics 20.

3. Although certain teachers are able to adjust their teaching methods to approximate theoretically polarized models of

instruction for the purposes of an experiment, they revert to methods they used previously more readily than other teachers.

4. The extent to which a teacher is able to adjust his teaching method to fit a particular model is not related to his sex.

5. Outside observers rate isolated lessons for methods. Students' opinions as indicated on inventories reflect what has gone on in the classroom for some time. Scores from the two sources agree to the extent that the teachers' methods are consistent between lessons.

II. IMPLICATIONS

There are some rather definite implications to be gathered from the above summary and conclusions. These can be classified according to implications for in-service programs, for further research, and for collecting data on teachers' methods.

In-Service Programs

Six of the seven participating teachers in this study were found to be teaching the two treatment classes during the experiment using essentially different methods. Four of the seven teachers were found to be using different methods of instruction after the experiment as the result of the in-service program. This lends support to the view that a teacher's repertoire of methods

may be enlarged through a methods in-service program. More in-service programs should be designed with the primary purpose of providing teachers with a variety of teaching methods. This may have significant implications for the problem of re-training of teachers who are now in the field and are not in a position to leave teaching for the purpose of re-vitalizing their methods of instruction.

Further Research

Williams (1966)[1] suggested that further research is needed to discover why certain teachers revert to their familiar methods more directly than others. This study found that certain teachers reverted more readily to methods of instruction familiar to them than other teachers. There is even more need for further research in the field of sustaining a variety of methods of instruction.

In contemplating a theoretical rationale for the flexibility in methods, the factors which effect changes in actual classroom teaching should be considered. It could be that because of particular circumstances, the teacher had to alter the teaching method to meet the goals of a particular lesson. There is need, therefore,

---

[1] J. D. Williams, "'Method-Reversion': The Problem of Sustaining Changes in Teacher-Behavior", Educational Research, Vol. VIII, (February, 1966), pp. 128-133.

for further research not only in the field of teacher training in flexibility of methods, but also in the area of methods necessary for achieving certain goals or objectives most effectively.

This study was based on one kind of in-service program. Further research needs to be carried out on the potential of a variety of in-service programs for effecting change in teachers instructional methods.

This study was not able to significantly relate flexibility of methods with certain selected characteristics although some direction was suggested. More research may bring out some definite patterns with respect to flexibility of methods and certain teacher characteristics.

Collecting Data on Teachers' Methods

Data from the two instruments of this study compared for teachers methods before the in-service treatment and after the experimental period. Data did not compare for flexibility of methods. This implies that when a teacher's methods of instruction are in a state of flux, students are not able to pick up the differences as clearly as an outside observer. It could be that the transition from one method to the other was so subtle that many of the students were not aware of it. More research is suggested on the extent to which students can identify changes in teaching methods.

Data gathered in this study was translated into ranks. Nonparametric statistical tests were carried out on the rank orders so obtained. Replications of this study are suggested in which random samples of mathematics teachers and different types of in-service programs are used. Replications of this study are also suggested at other grade levels and in other subject areas.

III. LIMITATIONS

This study was limited in several ways. First, because of limited resources, only one observer was able to rate teachers on the methods they used. The Observer Rating scores may have been flavoured with his biasness. Second, this study was limited by its developmental nature. Although the procedures and instruments were developed and tested before the in-service treatment, time prevented their refinement to the point that could have yielded more clear-cut results. Third, this study was limited by the number of subjects that were able to take part in the experiment. A random sample of High School mathematics teachers may have yielded different results.

BIBLIOGRAPHY

BIBLIOGRAPHY

Amidon, Edmund and Elizabeth Hunter. "Verbal Interaction in the Classroom - The Verbal Interaction Category System," Temple University, 1963. (unpublished manuscript).

Bales, Robert F. and Henry Gerbrands. "The 'Interaction Recorder,' An Apparatus and Check List for Sequential Content Analysis of Social Interaction," Human Relations, Vol. I, (November, 1948), pp. 456-463.

Beberman, Max. "Improving High School Mathematics Teaching," Educational Leadership, Vol. XVII, (december, 1959), pp. 162-166.

Bruner, Jerome S. Process of Education. Cambridge: Harvard University Press, 1960.

Bryan, Roy C. and Karl DuBois. "Comparison of Two Rating Scales Used to Obtain Student Reactions to Chemistry Teachers," School Science and Mathematics, Vol. LXI, (March, 1961), pp. 175-180.

Cogan, Morris L. "The Behavior of Teachers and the Productive Behavior of Their Pupils: 1. 'Perception' Analysis," Journal of Experimental Education, Vol. XXVII, (December, 1958), pp. 89-105.

Cook, Walter W. and Carroll H. Leeds. "The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes," Journal of Experimental Education, Vol. X, (December, 1947), pp. 149-159.

Cooksey, Henry B. "Evaluation of an In-Service Education Program for English Teachers," Unpublished Doctoral Dissertation, University of Texas, 1965.

Davis, Robert B. Discovery in Mathematics: A Text for Teachers, Palo Alto: Addison-Wesley, 1964.

______. Discovery in the Teaching of Mathematics. In Learning by Discovery: A Critical Appraisal, eds. L. S. Schulman and E. R. Kieslar, Chicago: Rand McNally and Company, 1966.

Denemark, George W. and James B. MacDonald. "Pre-Service and In-Service Education of Teachers," Review of Educational Research, Vol. XXXVII, (June, 1967), pp. 240-241.

Department of Extension, Continuing Professional Education for Teachers, Edmonton: University of Alberta, 1967.

Dienes, Zoltan P. Power of Mathematics, London, England: Hutchinson Educational Ltd., 1964.

Dossett, Marie J. "An Analysis of the Effectiveness of the Workshop as an In-Service Means for Improving Mathematical Understandings of Elementary School Teachers," Unpublished Doctoral Dissertation, University of Michigan State, 1964. (microfilmed.)

Fenton, Edwin and Richard B. Ford. "Organizing Summer Institutes for Teachers: Some Reflections," The Journal of Teacher Education, Vol. XVII, (Spring, 1966), pp. 53-59.

Ferguson, George A. Statistical Analysis in Psychology and Education, New York: McGraw-Hill Book Company, 1966.

Fischler, Abraham S. "Change In Classroom Behavior Resulting from an In-Service Program Utilizing Television," School Science and Mathematics, Vol. LXVII, (April, 1967), pp. 321-324.

Flanders, Ned A. "Analyzing Teacher Behavior," Educational Leadership, Vol. XIV, (December, 1961), pp. 173-180.

___ ___. "Teacher Behavior and In-Service Programs," Educational Leadership, Vol. XXI, (October, 1963), pp. 25-30.

_______. Teacher Influence, Pupil Attitudes, and Achievement In Studying Teaching, eds. James Raths et al. Englewood Cliffs: Prentice-Hall, 1967.

Guilford, Joy P. Fundamental Statistics in Psychology and Education, 4th edition, New York: McGraw-Hill Book Company, 1965.

Hawkins, David. Learning the Unteachable In Learning by Discovery: A Critical Appraisal, eds. L. S. Schulman and E. R. Kieslar, Chicago: Rand McNally and Company, 1966.

Heideman, Robert G. "National Science Foundation Academic Year Institutes for Secondary School Teachers of Science and Mathematics Held at the University of Wisconsin, 1956-1959: An Evaluation of the Background, Training, Placement, and Occupational Mobility of the Participants," Unpublished Doctoral Dissertation, University of Wisconsin, 1962. (microfilmed.)

Hilgert, Raymond L. "Teacher Reaction to Summer Workshops," School and Community, Vol. LIV, (January, 1968), pp. 14-15.

Houston, Robert W., Claude Boyd, and Vere M. De Vault. "An In-Service Program for Intermediate Grade Teachers," Arithmetic Teacher, Vol. VIII, (February, 1961), pp. 65-68.

Hughes, Marie. Development of the Means for the Assessment of the Quality of Teaching in Elementary Schools, Salt Lake City: University of Utah Press, 1959.

Jirik, Edward F. "A Statistical Analysis of 'Level of Achievement' as Related to Certain Supervisory Services and In-Service Participation," Unpublished Doctoral Dissertation, Ohio University, 1962.

Johnston, Ross. "The Mathematizing Mode," University of Alberta, Edmonton, 1968 (A working paper.)

Joyce, Bruce R. and Richard E. Hodges. "Instructional Flexibility Training," The Journal of Teacher Education, Vol. XVII, (Winter, 1966), pp. 409-416.

Kingston, Albert J. and George L. Newsome, "The Relationship of Two Measures of Authoritarianism to the Minnesota Teacher Attitude Inventory," Journal of Psychology, Vol. XVIX, (February, 1960), pp. 333-338.

MacLean, W. B. et al. Secondary School Mathematics Grade Eleven, Alberta edition, Vancouver: Copp-Clark Publishing Co. Ltd., 1966.

Medley, Donald M. "Experiences with the OScAR Technique," Journal of Teacher Education, Vol. XIV, (September, 1963), pp. 267-272.

______, and Harold E. Mitzel. Measuring Classroom Behavior by Systematic Observation. In Handbook of Research on Teaching, ed. Nathaniel L. Gage, Chicago: Rand McNally and Company, 1963.

Nelson, Lois N. "Teacher Leadership: An Empirical Approach to Analyzing Teacher Behavior in the Classroom," The Journal of Teacher Education, Vol. XVII, (Winter, 1966), pp. 217-425.

O'Rourke, Mary A. and William H. Burton. Workshop for Teachers, New York: Appleton-Century-Crofts, Inc., 1957.

Perkins, Hugh V. "A Procedure for Assessing the Classroom Behavior of Students and Teachers," American Educational Research Journal, Vol. I (November, 1964), pp. 249-260.

Polya, George. Mathematics Discovery: On Understanding, Learning and Teaching Problem Solving, Vol. 1, New York: John Wiley and Sons, 1962.

Roscoe, Douglas K. "An Analysis of Two Mathematics Workshops for Teachers and Outcomes as Reflected in Participating Schools," Unpublished Doctoral Dissertation, Alabama Polytechnic Institute, 1958. (microfilmed.)

Schueler, Herbert, Milton J. Gold, and Harold E. Mitzel. The Use of Television for Improving Teacher Training and for Improving Measures of Student-Teacher Performance, Phase I, New York: Hunter College, 1962.

Shaver, James P. A Study of Teaching Style: The Investigation Through Systematic Observation of the Ability of Experimental Teachers to Conform to Two Models of Teaching. Cambridge, Massachusetts: Harvard University Archives, 1961.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. Toronto: McGraw-Hill Book Company, 1956.

Smith, Othanel B. "A Concept of Teaching," Teachers College Record, Vol. LXI, (February, 1960), pp. 229-241.

Travers, Robert M. W. An Introduction to Educational Research, First edition, New York: The Macmillan Company, 1958.

______. An Introduction to Educational Research, Second edition, New York: The Macmillan Company, 1967.

Wallen, Norman E. and Robert M. W. Travers. Analysis and Investigation of Teaching Methods. In Handbook of Research in Teaching, editor Nathaniel L. Gage, Chicago: Rand McNally and Company, 1963.

Williams, J. D. "'Method-Reversion': The Problem of Sustaining Changes in Teacher-Behavior," Educational Research, Vol. VIII, (February, 1966), pp. 128-133.

Winer, B. J. Statistical Principles in Experimental Design, New York: McGraw-Hill Book Company, 1962.

Withall, John. "Conceptual Frameworks for the Analysis of Classroom Social Interactions," Journal of Experimental Education, Vol. XXX, (June, 1962), pp. 307-308.

Worthen, Blaine R. Discovery vs. Expository Sequencing in Six Weeks of Classroom Instruction in Mathematics. In Sequence Characteristics of Text Materials and Transfer of Learning, Authors Gabriel M. Della-Piana and others, Salt Lake City: University of Utah, 1965.

Wright, Muriel J. "Development of an Instrument for Studying Verbal Behaviors in a Secondary School Mathematics Classroom," Journal of Experimental Education, Vol. XXVIII, (December, 1959), pp. 103-121.

APPENDIXES

TEACHER CHARACTERISTICS QUESTIONNAIRE

May I impose on you to fill out the following questionnaire. The information will be treated in confidence and will be used in summary form in my study.

1. Name: ______________________________

   Home Address: ______________________________

   Home Phone: ____________________

2. Age: ____________________

   Years of teaching experience: __________

   Years of experience in teaching High School Mathematics: ______________

   Number of times you have taught the present High School Math 20 course: __________

3. Please list university courses you have in Math and year taken:

   Course Name and Number Year

   a)

   b)

   c)

   d)

   e)

   f)

   g)

   h)

University Degrees you have			
Year Obtained			
Institution			

4. First Class Time Taught __________ No. of Students __________

   Second Class Time Taught __________ No. of Students __________

POST - MEETING IN-SERVICE EVALUATION SHEET

Name: ______________________________ Number: ______________

Your assistance in the evaluation of the in-service meetings will help the directors to a more complete appreciation of the progress and effectiveness of the in-service programs. Circle your opinion please.

1. How do you feel this meeting was today?

VERY SATISFYING	SATISFACTORY	UNDECIDED	UNSATISFACTORY	VERY UNSATISFACTORY

2. Do you feel you were adequately prepared in advance to take part in the discussion after the lecture?

VERY WELL	ADEQUATELY	SOMEWHAT LACKING	LACKING	NOT AT ALL

3. To what extent were the things you personally hoped to get out of the lecture and discussion different from what the group was trying to accomplish?

IDENTICAL	FAIRLY SIMILAR	UNRELATED BUT NOT INCOMPATIBLE	SOMEWHAT DIFFERENT	COMPLETELY OPPOSED

4. To what extent were the lecture and discussion in your opinion adequate in breadth?

VERY ADEQUATE	SATISFACTORY	ACCEPTABLE	SOMEWHAT LACKING	INADEQUATE

5. To what extent were the lecture and discussion in your opinion adequate in depth?

VERY ADEQUATE	SATISFACTORY	ACCEPTABLE	SOMEWHAT LACKING	INADEQUATE

6. How fully do you think the members were in accord with what the director was trying to accomplish today?

COMPLETE ACCORD	GOOD MAJORITY	ABOUT HALF	LARGE MAJORITY	SMALL MINORITY IN ACCORD

7. In what ways did they differ?

a)	d)
b)	e)
c)	f)

8. Did you find yourself wanting to say things during the meeting that you didn't actually say?

   VERY FREQUENTLY FREQUENTLY FAIRLY OFTEN A FEW TIMES NOT AT ALL

9. Please list any reasons you did not contribute.

10. List questions that you would have liked answered.

11. Comments and suggestions for improvement.

POST - EXPERIMENT QUESTIONNAIRE

Now that some time has elapsed since the "experimental" portion of the project has been completed, I would like to impose on you to fill out this questionnaire. Would you please indicate your feelings towards certain aspects of the project by circling your choices.

A. In-Service Program

1. Do you feel there were a sufficient number of sessions for the program's purposes? YES NO If NO please comment ____________

   ____

2. Were the in-service sessions spaced adequately so that you were able to consider the hand-outs thoroughly between sessions?

   TOO MUCH TIME GOOD NOT ENOUGH TIME

3. What changes would you recommend for some future in-service?

   ____

   ____

   ____

B. Project As A Whole

1. Considering personal effort, class time, etc., do you feel the project was worthwhile to you personally? YES NO

2. Do you feel you are making use of the experience gained as a result of your participation? YES NO In what ways? ________

   ____

   ____

3. Would you participate in another experiment of a similar nature at some future time? YES NO

4. How do you personally feel about the merits of the Mathematizing Mode as compared with your personal methods? ________________

   ____

5. Do you feel that established procedures prior to the project had any serious effect on your success with the Mathematizing Mode? YES NO

6. What do you feel the students gained from participating in the experiment? ______________________________

   ____

Thank you again for your kind and sincere co-operation in this project.

OBSERVER RATING SCALE OF TEACHING BEHAVIOR

Teacher's Name: ____________________ Class: ______ Date: ______

		YES				NO
A.	Teacher Omniscience					
1.	Teacher acts as primary source of knowledge.	0	1	2	3	4
2.	Students depend on teacher to help them solve problems.	0	1	2	3	4
3.	In checking, he uses proper application of appropriate rules.	0	1	2	3	4
4.	Teacher solves problems directly without much help from the students.	0	1	2	3	4
5.	The teacher has responses corrected before proceeding.	0	1	2	3	4
B.	Introduction of Generalization					
1.	The teacher introduces a specific topic to the class.	0	1	2	3	4
2.	Rules are given before examples. (Ex. clarify rules)	0	1	2	3	4
3.	Teacher summarizes after each sub-topic. (e.g. problem)	0	1	2	3	4
4.	Teacher encourages students to hypothesize at solutions.	4	3	2	1	0
5.	Generalization is delayed until many students are aware of it.	4	3	2	1	0
C.	Control of Pupil Interaction					
1.	Students are encouraged to offer complete, correct solutions.	0	1	2	3	4
2.	Teacher questions students on direct items of the topic.	0	1	2	3	4
3.	Classroom interaction deviates from the topic.	4	3	2	1	0
4.	Students are encouraged to reveal rules that apply.	0	1	2	3	4
5.	Premature verbalization of the rule is discouraged.	4	3	2	1	0
D.	Method of Answering Questions					
1.	The teacher answers by reiterating and explaining the rule.	0	1	2	3	4
2.	The teacher gives examples to clarify application of rule.	0	1	2	3	4
3.	Teacher answers student's question rather than referring to class.	0	1	2	3	4
4.	The teacher answers by referring to student's computational sequence.	4	3	2	1	0
5.	The teacher uses sequenced examples without hinting the rule.	4	3	2	1	0
E.	Use of Student Responses					
1.	Students' suggestions are tested before comment or evaluation.	4	3	2	1	0
2.	Students evaluate each other's responses.	4	3	2	1	0
3.	Students' responses guide classroom interaction.	4	3	2	1	0
4.	Student responses bear directly on topic.	0	1	2	3	4
5.	Students give only possible hints to the solution.	4	3	2	1	0
F.	Method of Eliminating False Concepts					
1.	Students are warned of errors made in applying the rule.	0	1	2	3	4
2.	Students are warned of problems with special peculiarities.	0	1	2	3	4
3.	Students are cautioned about false generalizations.	0	1	2	3	4
4.	Students are given specific ways of working problems.	0	1	2	3	4
5.	Teacher leads class to overgeneralization of rule.	4	3	2	1	0

STUDENT INVENTORY OF TEACHER BEHAVIOR

Teacher: ______________________ Class: ________ Date: ________

Directions: We wish to know some of the kinds of things teachers and students do in mathematics classrooms. In this inventory, each of you will tell how often certain things are done in your math classroom. After your sheets have been scored, they will be destroyed. No one will know what answers you gave. There are no correct or wrong answers.

Read each of the items carefully and fill in the blank of the letter on the answer sheet that most accurately describes how often the event takes place in your math class.

1. When an answer is wrong, our teacher ____ tells us immediately.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

2. Our teacher ____ shows us how to solve typical problems.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

3. When we ask the teacher how to solve a problem, he ____ shows us.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

4. When we ask the teacher how to solve a problem, he ____ gives us only hints that we may use.

   A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

5. In taking up work in class, the teacher ____ sees that we get correct answers to all the problems and questions asked.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

6. Our teacher ____ explains each new rule before we are given examples to work out.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

7. We ____ work on a set of problems without being given any definite ways of working them out.

   A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

8. We ____ take up sample problems before we begin to work on a set of exercises.

   A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

9. Our teacher ____ encourages us to hypothesize or make guesses at solutions.

   A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

10. When we begin discussing work in class we ____ know what rules we will discuss.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

11. We seem ____ to be discussing more than one thing at a time in class.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

12. We are ____ encouraged to try solving problems even if our method may not work.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

13. Our teacher would ____ rather ask a question about the problem than give the correct answer.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

14. When one of us works problems out for other class members, our teacher is ____ unhappy.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

15. The teacher is ____ willing to discuss any math problems in class even if they are not on the topic.

    A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

16. In answering questions in class, our teacher ____ uses his own examples rather than ours.

    A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

17. In answering my questions, our teacher ____ refers back to the rules rather than what I did.

    A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

18. In our class, we are ____ given a direct answer to our question or problem.

    A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

19. Our teacher ____ gives us a rule to use for solving new kinds of problems.

    A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

20. When we ask questions in class, the teacher ____ would rather that someone from the class answer them.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

21. We ____ end up with two or three ways of solving the same type of problem.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

22. Our teacher ____ gives us a chance to try our own method before he points out his method.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

23. When a student gives a wrong answer, our teacher ____ appears unhappy.

A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

24. We are ____ asked to solve problems in only one way.

A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

25. I feel that I can solve problems ____ in any way that yields a correct solution.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

26. When a rule won't apply to all problems, our teacher ____ warns us when not to use it.

A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

27. In our assignments we are ____ given problems which cannot be solved using rules discussed in class.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

28. When a new problem comes up, our teacher ____ shows us exactly which rules to use.

A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

29. Our teacher ____ lets us try for ourselves to use a rule on a problem even when he knows it does not work.

A. Almost Never B. Seldom C. Sometimes D. Often E. Almost Always

30. We are ____ cautioned to think through our problems and solutions carefully.

A. Almost Always B. Often C. Sometimes D. Seldom E. Almost Never

APPENDIX D.1

AN EXAMPLE OF THE CALCULATION OF THE WILCOXON T STATISTIC*

Category	Mathematizing Class Scores	Expository Class Scores	d	Rank of d	Rank with less freq. sign
A	24	19	5	4.5	
B	34	32	2	3	
C	53	48	5	4.5	
D	41	34	7	6	
E	43	42	1	1.5	
F	32	33	-1	-1.5	1.5
					T= 1.5

*Based on pre-treatment scores from the Student Inventory for teacher 1.

APPENDIX D.2

TEACHER CATEGORY MEANS BASED ON PRE-TREATMENT DATA
FROM THE OBSERVER RATING SCALE

Category	Teacher						
	1	2	3	4	5	6	7
A	33	14	23	24	26	30	26
B	33	26	26	25	38	25	29
C	33	24	36	38	35	35	40
D	29	23	36	38	35	40	35
E	39	26	40	31	39	48	39
F	30	15	21	24	24	28	26

TEACHER CATEGORY MEANS BASED ON PRE-TREATMENT DATA
FROM THE STUDENT INVENTORY

Category	Teacher						
	1	2	3	4	5	6	7
A	27	17	21	21	24	23	21
B	33	25	27	26	32	27	28
C	51	46	50	45	50	38	44
D	38	33	43	39	41	35	39
E	43	49	55	50	70	59	52
F	33	21	21	30	24	24	26

APPENDIX E

CALCULATION OF FLEXIBILITY OF METHODS FOR TEACHERS FROM THE OBSERVER RATING SCALE

Teacher	PRE-TREATMENT		DURING-TREATMENT		Flexibility of Method (3-1)+(2-4)
	1 Math. Class Method Index	2 Exp. Class Method Index	3 Math. Class Method Index	4 Exp. Class Method Index	
1	30	34	62	35	31
2	20	22	63	36	29
3	30	32	66	37	31
4	28	31	48	43	8
5	32	31	67	38	28
6	35	34	59	40	18
7	34	32	63	41	20

CALCULATION OF FLEXIBILITY OF METHODS FOR TEACHERS FROM THE STUDENT INVENTORY

Teacher	PRE-TREATMENT		DURING-TREATMENT		Flexibility of Method (3-1)+(2-4)
	1 Math. Class Method Index	2 Exp. Class Method Index	3 Math. Class Method Index	4 Exp. Class Method Index	
1	38	36	60	36	22
2	33	30	70	33	34
3	36	36	65	40	25
4	34	33	45	44	0
5	39	41	65	45	22
6	35	33	65	38	25
7	36	33	62	35	24

APPENDIX F.1

PRE-TREATMENT SCORES OBTAINED BY TEACHERS ON OBSERVER RATING SCALE

	Teacher																											
	1				2				3				4				5				6				7			
	Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp	
Category	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	35	25	35	35	15	10	15	15	20	30	20	20	25	20	25	25	20	20	40	25	35	35	25	25	25	30	25	25
B	30	35	40	25	30	15	25	35	25	25	25	30	30	15	30	25	30	25	25	30	25	25	30	30	35	25	25	30
C	30	35	35	30	35	20	15	25	30	30	40	45	45	40	30	35	35	40	30	35	40	30	35	35	35	40	40	45
D	25	25	30	35	20	20	25	25	40	35	45	25	35	25	45	45	30	35	45	30	40	40	45	35	40	30	35	35
E	35	35	45	40	30	25	30	20	45	40	40	35	35	30	25	35	45	50	30	30	50	45	40	55	45	35	45	30
F	20	30	35	35	10	15	20	15	15	20	20	30	25	15	30	25	25	25	25	20	30	20	30	30	30	35	20	20
Method Index	29	31	35	33	23	18	22	23	29	30	32	31	33	24	31	32	31	33	33	28	37	33	33	35	35	33	32	31

APPENDIX F.2

PRE-TREATMENT SCORES OBTAINED BY TEACHERS ON STUDENT INVENTORY

Category	Teacher 1 Math	Teacher 1 Exp	Teacher 2 Math	Teacher 2 Exp	Teacher 3 Math	Teacher 3 Exp	Teacher 4 Math	Teacher 4 Exp	Teacher 5 Math	Teacher 5 Exp	Teacher 6 Math	Teacher 6 Exp	Teacher 7 Math	Teacher 7 Exp
A	24	29	19	14	20	21	18	23	22	26	25	20	21	20
B	34	32	25	24	24	30	28	24	28	36	26	27	29	29
C	53	48	45	46	46	54	50	39	52	47	42	34	43	45
D	41	34	33	33	47	38	41	37	38	43	36	34	43	34
E	43	42	50	47	59	51	48	52	69	71	60	58	53	50
F	32	33	24	18	20	21	19	21	24	24	22	26	30	21
Method Index	38	36	33	30	36	36	34	33	39	41	35	33	36	33

APPENDIX F.3

DURING TREATMENT SCORES OBTAINED BY TEACHERS ON OBSERVER RATING SCALE

	Teacher																											
	1				2				3				4				5				6				7			
	Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp	
Category	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	50	60	25	25	65	55	25	30	65	60	30	20	40	40	35	35	55	65	35	25	60	55	25	35	60	55	25	30
B	70	65	35	30	70	60	30	40	60	65	35	35	45	50	40	30	65	80	40	30	65	60	30	40	65	60	35	40
C	55	55	45	35	60	65	45	40	55	60	40	45	40	60	50	55	50	75	45	40	55	55	40	45	60	65	45	55
D	65	70	35	40	75	75	35	40	90	70	40	40	55	50	45	45	65	80	35	40	70	65	35	50	70	75	50	50
E	65	75	40	45	55	70	55	30	85	75	45	55	55	55	50	60	75	75	55	45	75	60	45	60	75	70	55	55
F	55	60	30	30	50	50	20	35	50	55	30	25	35	45	25	30	50	70	40	20	50	40	35	35	50	45	20	35
Method Index	60	64	35	34	63	63	35	36	68	64	37	37	48	60	41	44	60	74	42	33	63	56	35	44	63	62	38	44

APPENDIX F. 4

TEACHERS' DURING EXPERIMENT SCORES FROM STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Category	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	54	28	68	20	62	24	30	37	56	31	64	23	57	18
B	63	27	67	26	60	31	39	34	66	38	66	35	62	28
C	60	49	64	44	59	53	54	54	63	48	57	39	62	44
D	62	39	80	35	81	42	54	50	68	44	74	42	70	39
E	71	45	79	55	82	63	59	59	85	75	77	62	74	57
F	52	31	59	20	49	26	31	28	55	34	54	30	49	23
Method Index	60	36	66	33	66	42	45	44	66	45	65	39	62	35

APPENDIX F.5

POST -TREATMENT SCORES OBTAINED BY TEACHERS ON OBSERVER RATING SCALE

	Teacher																											
	1				2				3				4				5				6				7			
	Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp		Math		Exp	
Category	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	45	45	40	35	20	20	20	20	30	45	35	40	25	30	25	35	45	55	30	55	25	25	20	25	40	45	20	30
B	40	40	50	40	40	35	30	30	35	40	60	55	40	45	30	45	60	60	45	40	45	55	35	35	50	45	30	40
C	45	35	55	35	30	30	25	35	45	55	50	35	35	35	40	40	50	45	40	50	30	35	40	40	60	65	45	35
D	40	45	45	45	35	25	15	40	40	50	45	50	50	50	35	45	55	55	50	45	50	45	45	35	45	45	45	45
E	50	50	60	55	45	40	35	30	60	60	60	45	30	40	50	65	50	65	60	60	40	40	50	30	65	70	55	50
F	45	30	20	30	25	15	15	15	50	35	55	30	20	25	20	30	35	30	35	35	20	30	25	15	50	40	30	30
Method Index	44	41	45	40	33	34	23	28	43	48	51	43	33	38	33	43	49	52	43	48	35	38	36	30	52	52	38	38

APPENDIX F.6

TEACHERS' POST-EXPERIMENT SCORES FROM STUDENT INVENTORY

	Teacher													
	1		2		3		4		5		6		7	
Category	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp	Math	Exp
A	29	27	22	18	43	24	26	36	38	31	33	28	33	20
B	33	28	23	24	48	37	29	35	44	39	34	32	41	31
C	53	45	42	42	51	46	51	55	56	42	46	43	51	42
D	48	42	36	35	63	40	46	53	50	41	40	37	52	34
E	58	49	47	50	69	59	61	60	71	68	70	67	65	53
F	32	28	27	17	32	24	24	33	40	33	28	32	35	26
Method Index	42	36	33	31	51	39	40	45	50	42	42	40	46	34

APPENDIX G

SUMMARY OF THE COMMENTS FROM THE POST-EXPERIMENT QUESTIONNAIRE

The overall experiment involved the time and effort of the subjects over a period of five months. This investigator was interested in the teachers' reactions to the project.

Below is a summary of the results of the Post-Experiment Questionnaire.

A. <u>In-Service Program</u>

1. All teachers agreed that there were a sufficient number of in-service sessions for the program's purposes.

2. The response of all teachers was GOOD to the second question.

3. Some changes recommended were:

   a. Mathematizing activities be developed for the complete course.

   b. Teachers have the opportunity to observe demonstration classes using the Mathematizing mode.

B. <u>Project as a Whole</u>

1. All teachers felt the project was worthwhile from their point of view.

2. All but two teachers felt that they were personally making

use of the experience gained in their instruction. Two teachers felt they did not have sufficient time to prepare Mathematizing activities after the experimental period.

3. All teachers expressed a willingness to participate in future projects.

4. Some teachers felt that Mathematizing method of teaching took too much class time to cover the same units. Others felt that the new experiences gained by the students and teachers were worth the extra time.

5. All teachers felt that the methods they used before the treatment had no serious effect on their ability with the Mathematizing method.

6. A variety of comments were given for this item. Some teachers felt the students gained a deeper insight into mathematics when taught with the Mathematizing method. Others noticed an appreciable increase in the students' enthusiasm for mathematics. Still others felt that the students gained from the increased classroom interaction in the mathematizing classes.